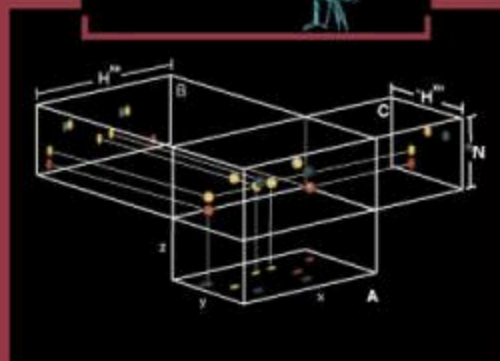
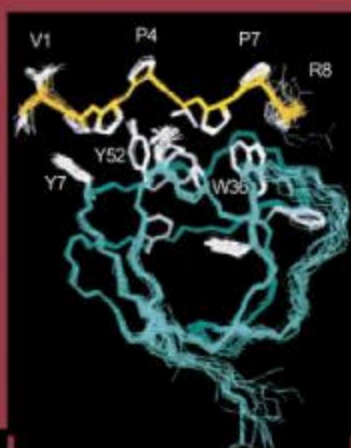


# NMR SPECTROSCOPY AND ITS APPLICATION TO BIOMEDICAL RESEARCH

Edited by  
**SUSANTA K. SARKAR**



**ELSEVIER**

## Preface

NMR has become the most diverse spectroscopic tool available to date in biomedical research. It is now routinely used to study biomolecular structure and dynamics particularly as a result of the recent developments of a cascade of highly sophisticated multidimensional pulse sequences, and of advances in genetic engineering to produce biomolecules, uniformly or selectively enriched with  $^{13}\text{C}$ ,  $^{15}\text{N}$  and  $^2\text{H}$ . This book, written by acknowledged experts, provides an up-to-date treatment of the current status of multi-dimensional NMR, including the basic aspects, and its application to problems of biomedical interest.

William Westler, in Chapter 1, provides a practical introduction to two-dimensional NMR through coherent flow network. This description makes it easy to conceptualize the pulse sequence details and provides a basis for understanding further complicated sequences. In Chapter 2, Luciano Mueller and N. Vasant Kumar describe the current status of multi-dimensional NMR and its utility in structure determination of proteins and nucleic acids. In addition to a brief description of different classes of multidimensional experiments they also summarize the spectrometer requirements for implementing these sequences.

Because of the recent developments in molecular biology techniques for preparing labeled samples, it is now possible to perform NMR studies on large macromolecules. An overview of different procedures for isotopic enrichment of proteins by labeled amino acids is given by Brian Stockman in Chapter 3. A critical discussion of the advantages and disadvantages of different procedures is also provided. Chapter 4, by Paul Weber, is a critical survey of a number of methods used for structure calculation of proteins from NMR data and provides the user with an insight into the process. This chapter also provides a list of different options available for structure calculation.

In order to understand protein function, it is essential to consider the internal mobility of proteins, which can be studied by recently introduced methods to measure heteronuclear relaxation. Linda Nicholson, Lewis Kay and Dennis Torchia give an introduction to relaxation theory including the recently developed pulse sequences to measure relaxation parameters in Chapter 5. A discussion of the data processing steps and a critical analysis of the motional parameters are also included. The

entire process is demonstrated with examples of detailed dynamic studies on staphylococcal nuclease.

Chapters 6 and 7 review the basic NMR methods and their modifications, particularly useful for resonance assignments in nucleic acids and carbohydrates. David Wemmer, in Chapter 6, discusses the steps necessary for building solution structure of nucleic acids from the available NMR data and the NMR methods for analyzing dynamics of nucleic acids. Specific examples are given to illustrate these approaches and a critical analysis is given about the specific issues for structural studies of DNA, RNA and their complexes with drugs and proteins. In Chapter 7, Laura Lerner describes the NMR methods and their application to structural analysis of oligosaccharides and their interactions with receptors.

Solid state NMR provides a powerful tool for structural and dynamic studies of many biological molecules not amenable to (a) solution NMR studies because of their size or (b) X-ray diffraction because of the unavailability of single crystal samples. Alexandra Simmons, Susanta Sarkar and Lynn Jelinski, in Chapter 8, outline the differences between nuclear interactions in solution and solid state and review the techniques commonly used in solid state NMR to obtain high resolution spectra from solid samples. Selected examples from the literature are used to demonstrate the application of solid state NMR studies to questions of biomedical interest.

In summary, it is hoped that this book would be useful to NMR spectroscopists, chemists, biochemists, and to molecular biologists interested in the use of NMR techniques for solving biological problems. The intended audience is NMR spectroscopists who are interested in biological problems and biologists who would like to use NMR.

I would like to thank all the authors for their contributions, which reflect the opportunities and the challenges of NMR spectroscopy and its application to biomedical research. I would also like to thank Mrs. Marjorie J. Krog for her help in preparing this book.

Susanta K. Sarkar  
*King of Prussia, PA*  
*October 1996*

**List of contributors**

*Lynn W. Jelinski*, Center for Advanced Technology in Biotechnology, Cornell University, 130 Biotechnology Building, Ithaca, NY 14853-2703, USA

*Lewis E. Kay*, Department of Medical Genetics, Biochemistry and Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A8

*N. Vasant Kumar*, Rhone-Poulenc Rorer, SW-3, Medicinal Chemistry, 500 Arcola Road, Collegeville, PA 19426, USA

*Laura E. Lerner*, Analytical Chemistry, MS 62, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080, USA

*Luciano Mueller*, Bristol-Myers Squibb, Pharmaceutical Research Institute, Princeton, NJ 08543, USA

*Linda K. Nicholson*, Section of Biochemistry, Molecular and Cellular Biology, Cornell University, Ithaca, NY 14853, USA

*Susanta K. Sarkar*, Department of Physical and Structural Chemistry, SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, PA 19406, USA

*Alexandra Simmons*, Research and Development Center, DuPont Canada, Inc., Kingston, Ontario, Canada K7L 5A5

*Brian J. Stockman*, Physical and Analytical Chemistry, Upjohn Laboratories, M/S 7255-209-0, The Upjohn Company, 301 Henrietta Street, Kalamazoo, MI 49001, USA

*Dennis A. Torchia*, Bone Research Branch, Protein Biophysics Section, National Institute of Dental Research, National Institutes of Health, Building 30, Room 106, Bethesda, MD 20892, USA

*Paul L. Weber*, Tripos Associates, 1699 South Hanley Road, St. Louis, MO 63144, USA



*David E. Wemmer*, Department of Chemistry, University of California,  
Structural Biology Division, Lawrence Berkeley Laboratory, Berkeley,  
CA 94720, USA

*William M. Westler*, National Magnetic Resonance Facility at Madison  
(NMRFAM), Department of Biochemistry, College of Agricultural  
and Life Sciences, University of Wisconsin-Madison, 420 Henry Mall,  
Madison, WI 53706, USA

## Chapter 1

# Two-Dimensional NMR Spectroscopy: A Graphical, “Top-Down” Description

WILLIAM M. WESTLER

### ABSTRACT

An introduction to the principles and uses of homonuclear and heteronuclear two-dimensional NMR spectroscopy is presented. The pulse sequences and phase cycling procedures for a few of the most common experiments are analyzed in detail. The experiments are described by using a visual description of spin system evolution and coherence transfer processes. This method of description allows for the discussion of any multidimensional, multinuclear pulse sequence by a “top-down” approach.

### 1. INTRODUCTION

Two-dimensional NMR spectroscopy has become a mainstay experiment in many disciplines [1]. The use of two- and higher- dimensional NMR has greatly advanced the applications of NMR to biochemical systems. Before the introduction of two-dimensional spectroscopy, researchers were generally limited to assigning the few resonances that are resolved in one-dimensional spectra or to use specific  $^{13}\text{C}$  or  $^{15}\text{N}$  labels. Over the past decade, techniques have been developed for the assignment of most, if not all, of the NMR resonances in a macromolecule, while advances in the field of molecular biology have led to the production of the necessary isotopically labeled macromolecules. The assignment of individual resonances to particular nuclei in the molecules

provides a multitude of probes with which to interrogate the molecular system. With the complete assignment of the proton network, the nuclear Overhauser effect yields distances between protons in macromolecules and, by a variety of methods, three-dimensional structures of macromolecules in solution can be obtained.

The concept of using more than one dimension in NMR spectroscopy was first introduced by Jeener [2] and developed by Ernst and coworkers in the late 1970s [3]. Two-dimensional NMR takes advantage of the non-linear properties of the nuclear spin system by passing frequency, amplitude, and phase information from one nucleus to another. The transferred information is observed indirectly as a modulation of the detected nuclei. The mechanisms of transfer can be classified as incoherent or coherent. The incoherent mechanism of information transfer uses either the dipolar interaction or physical chemical exchange, whereas the coherent mechanism relies on the information being passed through the scalar coupling interaction. Most multidimensional experiments are of the coherence transfer type. The information that is gained from these experiments is used to connect nuclei that are part of a scalar-coupled network of spins. Since the scalar coupling interaction occurs between nuclei that are one to a few chemical bonds apart, these experiments are used to obtain information about the primary structure of the molecule. The coupling constant information can be used to determine various dihedral angles within the scalar-coupled network and lends information about molecular secondary structure. While the number of experiments that use incoherent transfer of magnetization is small, these experiments hold a very important role in the determination of molecular structure. From the dipolar interaction, through-space distance information is obtained. The measurable distances are generally less than about 5 Å; since the information does not rely on the presence of a chemical bond between the interacting nuclei, primary, secondary and tertiary structural information is available.

A number of books [1,4] and review articles [5–7] have been published on the principles and uses of 2D NMR spectroscopy and I will not repeat the many references contained within those articles. My goal here is to present a practical introduction to the principles and uses of a few common 2D experiments. As an introduction to two-dimensional NMR methods, a simple, although naive, experiment is described that shows the fundamentals of multidimensional NMR by describing the behavior of nuclei undergoing exchange from one site to another in an idealized molecule. An introduction to heteronuclear coherence transfer

will lead into the common homonuclear and heteronuclear two-dimensional methods. Along with each experiment, an analysis of the pulse sequence and a phase cycling procedure is presented. At the end of each experiment, variants of the experiment and notes are presented with the interest mainly focused on improvements in the collection or processing of the data.

The next section introduces a graphical nomenclature for pulse sequences, the coherence flow network (CFN). The CFN represents the flow of coherence in multipulsed NMR experiments. This device, which is directly obtained from the product operator description [8], provides a simple visualization of multipulsed NMR experiments. Hopefully, the CFN approach will help bridge the gap between a qualitative description of the experiment and the detailed mechanism. Following that section, a detailed discussion of the mechanism and procedures of phase cycling is presented. Again, I will be verbose in the hope that the detail will help in understanding the apparent “magic” of this procedure.

## 2. COHERENCE FLOW NETWORKS

A popular approach to the description of NMR pulse sequences comes from a simple vector model [9,10] in which the motion of the spins subjected to RF pulses and chemical shifts is described by the rotations of a classical vector in three-dimensional space. The major weakness in this model is that spin systems that contain scalar coupling are not adequately described. The ultimate description of NMR experiments is through the application of density matrix theory. This approach, while extremely powerful, is often difficult to the uninitiated. In 1983, several groups independently introduced the product operator formalism as a simplification of density matrix theory [11–13]. In this treatment, the density matrix is expanded as a linear combination of a basis set of operators. The product operators resolve the density matrix into a set of elements that can be visualized as generalized magnetization vectors, referred to as coherences. The motion of the product operators subjected to RF pulses, chemical shifts, and coupling is analogous to the motion of the classical magnetization vectors in three-dimensional space and thus the popular vector model can be extended to spin systems that contain scalar coupling.

Currently, the majority of experiments in the literature are described by using one or another implementation of the product operator

formalism. Product operators, while very convenient for a simple description of NMR experiments, are not very compact for complex calculations. A given pulse sequence can generate many product operator terms that are multiplied by various cosine and sine coefficients. Many of these terms are not physically observable or are eliminated by phase cycling routines, so the description of pulse sequences with product operators is commonly abbreviated by retaining only the relevant terms. Usually, only a few of the important coherences are described at various points in the pulse sequence to give a sense of the flow of coherence from one section of the experiment to the next. Much of the detail of the mechanism of coherence transfer and modulation is removed, leaving behind the essentials of the pulse sequence. Eggenberger and Bodenhausen [14] have used a graphical approach to simplify the product operator notation. Their description (EB) does not compromise the rigor of the product operator treatment and allows for the exact description of experiments. While EB is a useful simplification for the practising NMR spectroscopist, it lacks an overview quality that can be useful in a “top-down” approach to NMR spectroscopy.

Introduced here is a systematic, graphical description of multipulse experiments in terms of the flow of coherence during a pulse sequence. Coherence flow network (CFN) diagrams can be used to accurately describe the mechanisms of coherence transfer and modulation used in multipulsed NMR experiments. Experiments can be written at a very qualitative level for “back-of-the-envelope” discussions or, with increasing detail, at a level that is equivalent to the product operator formalism. The “top-down” approach of CFN is a very natural procedure for the solution of problems involving NMR spectroscopy.

## **2.1. Description of the coherence flow network**

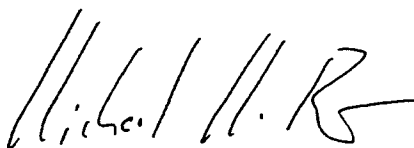
A coherence flow network (CFN) describes the flow of coherence during an NMR pulse sequence. The flow of coherence is the pathway followed by the signal of interest that is selected by the pulse sequence and its associated phase cycle. The CFN description does not detail the pulse sequence nor the phase cycle that is used to accumulate the experimental data, but the essence of the product operator description is abstracted into a visual form.

As an introduction to the uses of a CFN in the understanding of a complicated pulse sequence, the *three-dimensional* HNC0 experiment [15] is shown in Fig. 1.1. The spectrum obtained from this sequence

## Foreword

What would be molecular biology without NMR, and what would be NMR without its applications to molecular biology? — The world of science would miss a fruitful and exciting field. Next to clinical medicine, molecular biology has become the field most dependent on the recent progress in NMR technology. A powerful arsenal of versatile tools is available today for studying biomolecular structure and intramolecular dynamics. In particular, multi-dimensional spectroscopy has expanded in an unprecedented manner the possibilities of gaining insight into biological macromolecules.

This volume provides an up-to-date treatment of NMR methodology in view of biomedical research. Basic aspects as well as most refined modern pulse techniques are covered by a group of leading NMR spectroscopists. The book will undoubtedly prove useful in the hands of practising spectroscopists and biochemists who intend to apply modern NMR techniques.

A handwritten signature in black ink, appearing to read 'R. Ernst', with a stylized flourish at the end.

Richard R. Ernst  
*Zürich*  
October 1996

correlates the chemical shifts of the amide  $^1\text{H}$ , the amide  $^{15}\text{N}$ , and the attached carbonyl  $^{13}\text{C}$ . At the top of the figure is the pulse sequence that might be presented in a paper discussing this experiment. In such a paper [16], one might read "...magnetization originating from NH protons is transferred to the directly bonded  $^{15}\text{N}$  spins using an INEPT sequence, following which  $^{15}\text{N}$  magnetization evolves exclusively under the influence of the  $^{15}\text{N}$  chemical shift because of  $^1\text{H}$ ,  $^{13}\text{CO}$  and  $^{13}\text{C}^\alpha$  decoupling through the application of  $180^\circ$  pulses at the midpoint of the  $t_1$  interval...". The text continues to describe what happens next in the sequence continuing through the detection of the NH protons. This description is a detailed visualization of the experiment without any physical "picture" being drawn. At the bottom of Fig. 1.1 is the CFN of this experiment. The verbal description of the CFN is very similar to that above, but the advantage of the CFN to the reader is that no "minds eye" view of the coherence flow is necessary. The CFN is an attempt to graphically represent the verbal description accompanying a pulse sequence.

The basic rules for a CFN are:

- (1) A level is present for every spin involved in the experiment.
- (2) The coherence that is selected by the phase cycle and/or pulsed field gradients is followed over a path through the various states of the spin system (A, B, C, etc. in Fig. 1.1). This path describes chemical shift evolution, spin echos, anti-phase build-up, coherence transfer, etc.

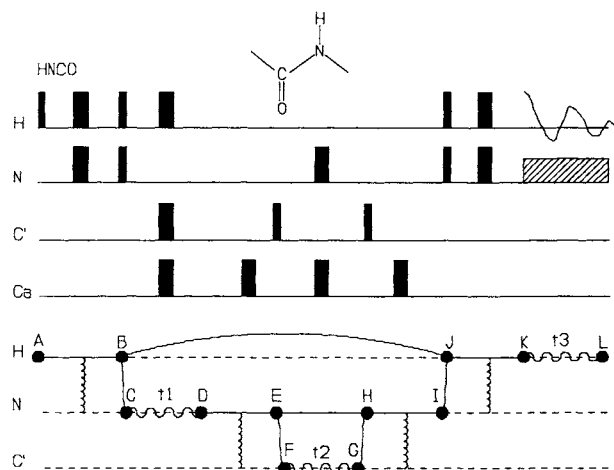


Fig. 1.1. Pulse sequence and CFN for the three-dimensional HNCO experiment.

(3) Various modifiers attached to the pathways represent active or passive coupling, or decoupling.

Each observed signal in a multipulsed NMR experiment can be systematically traced from a *source* of magnetization by *evolution* of coherence through intervening *states* to a final *sink* corresponding to the detection of the observed resonance. The evolution of coherence through the states represents the motion of the spin system under the influence of the Hamiltonian operators that are active during various evolution periods of the pulse sequence. By following the evolution of the spin coherence through the pulse sequence, the point-of-origin of information, such as the evolution period for indirectly detected nuclei and coherence transfer steps, is easily localized to a particular part of an experiment.



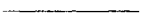
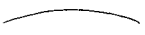
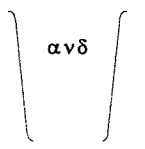
In a coherence flow network, the evolution of a coherence is assigned a primitive, graphical operator (*flow primitive*) between every pair of states that are visited by the coherence flow. Flow primitives have associated *modifiers* that detail the nature of interactions of the spin assigned to the flow primitive with other spins in the system. Flow primitives and modifiers are further combined systematically to construct a coherence flow network.

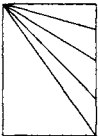




Table 1.1 contains descriptions of the primitive operators that are used in creating CFN diagrams. Every spin that is involved in the coherence flow is assigned a *tier* that is designated by a horizontal dashed line. On each spin tier, various CFN primitives are used to describe the evolution of the spin between states. States, designated by circles, are found at the junction of two adjacent evolutions. In an experiment with many different source spins, the coherence flow networks of the various spins overlap, possibly with multiple networks having a common sink. Similarly, diagrams with a single source spin that split into several paths having different sinks represent overlapping networks. However, for a given, observed coherence there is only one source and one sink.





If a spin is not involved in a coherence, i.e., the spin is at Boltzmann equilibrium or saturated, or if a coherence does not contribute to the final signal then there is no CFN flow primitive or modifier drawn on that spin tier during that period of the pulse sequence. CFN primitives that connect two states located on the same tier represent evolutions involving either transverse magnetization (e.g.,  $\mathbf{I}_x$ ), the Z component of an antiphase coherence (e.g.,  $2\mathbf{I}_z\mathbf{S}_x$ ), or a longitudinal spin order state (e.g.,  $2\mathbf{I}_z\mathbf{S}_z$ ) for the spin on that tier. Flow primitives drawn as either



TABLE 1.1  
Coherence flow network primitives

	Operator	Effect	Example
CS		Transverse magnetization with chemical shift precession.	$\mathbf{I}_x$ $=\omega_1 t \hat{\mathbf{I}}_x \Rightarrow$ $\mathbf{I}_x \cos(\omega_1 t) + \mathbf{I}_y \sin(\omega_1 t)$
TC		Transverse magnetization with constant time chemical shift precession.	$\mathbf{I}_x$ $=\omega_1 (T-t/2) \hat{\mathbf{I}}_x \Rightarrow$ $=\pi \hat{\mathbf{I}}_x \Rightarrow$ $=\omega_1 (T+t/2) \hat{\mathbf{I}}_x \Rightarrow$ $\mathbf{I}_x \cos(\omega_1 t) + \mathbf{I}_y \sin(\omega_1 t)$
E		Transverse magnetization without chemical shift precession.	$\mathbf{I}_y$ $=\omega_1 t \hat{\mathbf{I}}_x \Rightarrow$ $=\pi \hat{\mathbf{I}}_x \Rightarrow$ $=\omega_1 t \hat{\mathbf{I}}_x \Rightarrow$ $-\mathbf{I}_y$
Z		Evolution with retention of antiphase state.	$2 \mathbf{I}_x \mathbf{S}_x$ $=\pi/2 \ 2 \hat{\mathbf{S}}_x \mathbf{T}_x \Rightarrow$ $4 \mathbf{I}_x \mathbf{S}_y \mathbf{T}_x$
CT		Coherence transfer (between spins).	$2 \mathbf{I}_y \mathbf{S}_x$ $=\pi/2 (\hat{\mathbf{I}}_x + \hat{\mathbf{S}}_x) \Rightarrow$ $-2 \mathbf{I}_x \mathbf{S}_y$

	Operator	Effect	Example
T		Isotropic mixing	TOCSY, HOHAHA
N		Incoherent transfer by longitudinal cross-relaxation.	NOESY, Chemical Exchange
R		Incoherent transfer by transverse cross-relaxation.	ROESY, Chemical Exchange
A		Active coupling, oscillation between inphase and antiphase states.	$\mathbf{I}_x$ $= \pi J_{IS} t \, 2\hat{\mathbf{I}}_x \hat{\mathbf{S}}_z \Rightarrow$ $2\mathbf{I}_y \mathbf{S}_z \sin(\pi J_{IS} t)$
P		Passive coupling, no oscillation between inphase and antiphase states.	$\mathbf{I}_x$ $= \pi J_{IS} t \, 2\hat{\mathbf{I}}_x \hat{\mathbf{S}}_z \Rightarrow$ $\mathbf{I}_x \cos(\pi J_{IS} t)$

	Operator	Effect	Example
CC		Antiphase buildup due to cross correlation	$I_x$ $=\Gamma_{IS} * t \Rightarrow$ $a * 2I_x S_z$
T1		Longitudinal relaxation, $T_1$	Inversion-recovery
T1 R		Rotating frame longitudinal relaxation, $T_{1\rho}$	Spin-lock
T2		Transverse relaxation, $T_2$	Spin-echo

straight lines (Table 1.1 E) or wavy lines between states (Table 1.1 CS) represent inphase, transverse magnetization while a curved arc represents the Z state of an antiphase coherence (Table 1.1 Z). Flow primitives can have added modifiers that connect two different spin tiers. These primitives are vertical wavy lines representing “active” coupling (Table 1.1 A) or hashed boxes representing “passive” coupling (Table 1.1 P)

Active coupling causes the initial state to oscillate between inphase and antiphase states. If the initial state was inphase, then the final state is an antiphase state (with a coefficient of  $\sin(\pi J_{\text{IS}}t)$ ). Likewise, if the initial state was antiphase, then the final state is inphase. A passive coupling modulates an initial state by  $\cos(\pi J_{\text{IS}}t)$  but does not allow for the conversion between inphase and antiphase states. Couplings between more than one spin pair are designated by multiple **A** or **P** modifiers emanating from a single flow primitive.

CFN flow primitives connecting states on different tiers represent the transfer of coherence from one spin to another (Table 1.1 **CT**). These transfers usually pass through an antiphase state formed from the spin states involved in the coherence transfer. The **CT** primitive is represented by a solid line connecting the two states that are involved. Transfer by an isotropic mixing sequence (as in TOCSY or HOHAHA experiments) is represented by multiple **CT** primitives connecting one state to all other states that are involved in the isotropic mixing and then enclosed in a box (Table 1.1 **T**).

A CFN for an experiment is created by the combination of flow primitives, modifiers, and states. One begins by generating a tier for every spin that is involved in the coherence flow. The source and sink are chosen. Then flow primitives and modifiers are introduced to emulate the coherence flow between states during the pulse sequence. The CFN, which is generated, represents the overall flow of coherence *after all phase cycling and pulsed field gradient coherence selection*.

## 2.2. Composite rotations

The CFN representation is closely related to concept of composite rotations [1], in which several rotation operators are combined into one or a few simple rotations. Composite rotations can be used as “black boxes” for the description of common segments of pulse sequences. The evolution of coherence between states during a segment of a pulse sequence is dictated by the effective Hamiltonian operator that governs evolution during that segment. The effective Hamiltonian can consist of any combination of operators, such as, chemical shift operators, scalar coupling operators, dipolar operators, and radiofrequency fields. The evolution of the spin system under a given effective Hamiltonian operator can be written as the product of all of the unitary operators that represent the individual parts of the Hamiltonian. A series of  $i$  rotations  $U_i$  representing the various Hamiltonian operators during the pulse sequence can be written as:

$$\sigma_0 = U_1 \Rightarrow U_2 \Rightarrow U_3 \Rightarrow U_4 \Rightarrow \dots \Rightarrow U_i \Rightarrow \sigma_i \quad (2.2.1)$$

These distinct rotations can always be combined into a single total rotation applied to  $\sigma_0$ .

$$\sigma_0 = U_{\text{total}} \Rightarrow \sigma_i \quad (2.2.2)$$

Usually,  $U_{\text{total}}$  is a simple rotation or at most a short series of a few simple rotations. By reducing the pulse sequence to a series of simple propagators, the experimental method can be described in a simple and straightforward manner.

### 2.3. Spin echo sequence for an isolated spin

A spin echo sequence ( $\tau - 180_\phi - \tau$ ) refocuses the chemical shift (and any dephasing due to field inhomogeneity) of a transverse isolated spin, creating an effective Hamiltonian operator that does not contain a chemical shift operator. This sequence does not alter the state of the spin system except for a phase shift. The sequence of the rotation operators is

$$= \omega_1 t \hat{I}_z \Rightarrow = \pi \hat{I}_\phi \Rightarrow = \omega_1 t \hat{I}_z \Rightarrow \quad (2.3.1)$$

representing a free evolution under the chemical shift operator ( $\hat{I}_z$ ), a  $180^\circ$  pulse about a  $\phi$  axis, and finally, a second free evolution due to chemical shift. This sequence can be simplified by introducing a unit operator

$$= \pi \hat{I}_\phi \Rightarrow = -\pi \hat{I}_\phi \Rightarrow \quad (2.3.2)$$

at the beginning of sequence (2.3.1)

$$= \pi \hat{I}_\phi \Rightarrow \{ = -\pi \hat{I}_\phi \Rightarrow = \omega_1 t \hat{I}_z \Rightarrow = \pi \hat{I}_\phi \Rightarrow \} = \omega_1 t \hat{I}_z \Rightarrow. \quad (2.3.3)$$

The portion of sequence (2.3.3) isolated by brackets can be simplified by use of the equivalent rotation [8,11]

$$= -\pi \hat{I}_\phi \Rightarrow = \omega_1 t \hat{I}_z \Rightarrow = \pi \hat{I}_\phi \Rightarrow = -\omega_1 t \hat{I}_z \Rightarrow. \quad (2.3.4)$$

This and similar equivalent rotations are simple to remember:

*If there are two identical operators with opposite signs surrounding another operator, then the equivalent rotation is obtained by the rotation of the central operator by the third operator. This does not always give simple results, but with judicious choices of the inserted unit operators, the sequence can be often simplified.*

Upon substitution of the equivalent rotation into (2.3.3), the sequence becomes

$$=\pi\hat{\mathbf{I}}_y=>=-\omega_1t\hat{\mathbf{I}}_z=>=\omega_1t\hat{\mathbf{I}}_z=>. \quad (2.3.5)$$

The adjacent, inverse rotations  $=-\omega_1t\hat{\mathbf{I}}_z=>$  and  $=\omega_1t\hat{\mathbf{I}}_z=>$  are equivalent to the identity operator  $=\hat{\mathbf{I}}_z=>$ , which does not rotate the spin. The sequence can then be further simplified to

$$=\pi\hat{\mathbf{I}}_y=>. \quad (2.3.6)$$

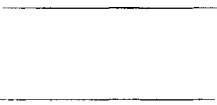



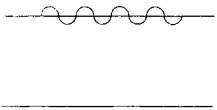

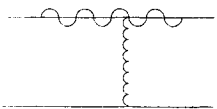

The final phase of the spin depends on the phase of the  $180^\circ$  pulse, but no chemical shift evolution occurs. This is also formally equivalent to the refocusing of spin isochromats in an inhomogeneous magnetic field where the chemical shift operator is replaced by a spatially-varying magnetic field strength.

In CFN notation this spin echo sequence is represented by a primitive that is diagrammed as a straight line (Table 1.1 E), signifying that no chemical shift evolution occurs between the two states that are connected by the spin echo sequence. The phase of the coherence is not explicitly represented in the CFN description. The phase information, however, is implicitly held in each individual state. By explicitly describing the states, CFN becomes equivalent to the product operator formalism.

#### **2.4. Coupled two-spin system — selective $I$ $180^\circ$ pulse**

Four important evolutions for coupled, two-spin systems and their CFN equivalents are depicted in Table 2.1. The spin echo sequence described in Section 2.3 was for an isolated spin; however, the same flow primitive is used in coupled spin systems to represent the evolution of a spin whose chemical shift is refocused and spin partners are decoupled (Table 2.1A). This pulse sequence consists of a  $180^\circ$  pulse on the transverse spin placed in the center of the evolution period with no pulse on the coupled partner. The rotation operators for this sequence are:

TABLE 2.1  
Several important CFN diagrams

CFN diagram	Pulse sequence	Composite rotation
<div>A</div> 		$=\pi\hat{\mathbf{I}}_{\phi}=>$
		$=\pi\hat{\mathbf{I}}_{\phi}=>$ $=\pi J_{IS}t2\hat{\mathbf{I}}_z\hat{\mathbf{S}}_z=>$
		$=\omega_I t\hat{\mathbf{I}}_z=>$
<div>D</div> 		$=\omega_I t\hat{\mathbf{I}}_z=>$ $=\pi J_{IS}t2\hat{\mathbf{I}}_z\hat{\mathbf{S}}_z=>$

$$=\omega_1 t \hat{\mathbf{I}}_z => =\pi J_{\text{IS}} t 2 \hat{\mathbf{I}}_z \hat{\mathbf{S}}_z => =\pi \hat{\mathbf{I}}_\Phi => =\pi J_{\text{IS}} t 2 \hat{\mathbf{I}}_z \hat{\mathbf{S}}_z => =\omega_1 t \hat{\mathbf{I}}_z => \quad (2.4.1)$$

The  $2 \hat{\mathbf{I}}_z \hat{\mathbf{S}}_z$  operator is the weak-scalar-coupling operator. Inserting the identity rotation,  $=\pi \hat{\mathbf{I}}_\Phi => =-\pi \hat{\mathbf{I}}_\Phi =>$ , between the first two operators in Eq. (2.4.1), one obtains

$$=\omega_1 t \hat{\mathbf{I}}_z => =\pi \hat{\mathbf{I}}_\Phi => \{-\pi \hat{\mathbf{I}}_\Phi => =\pi J_{\text{IS}} t 2 \hat{\mathbf{I}}_z \hat{\mathbf{S}}_z => =\pi \hat{\mathbf{I}}_\Phi =>\} =\pi J_{\text{IS}} t 2 \hat{\mathbf{I}}_z \hat{\mathbf{S}}_z => =\omega_1 t \hat{\mathbf{I}}_z =>. \quad (2.4.2)$$

The bracketed operators simplify to

$$=-\pi J_{\text{IS}} t 2 \hat{\mathbf{I}}_z \hat{\mathbf{S}}_z =>. \quad (2.4.3)$$

Sequence (2.4.3) combined with the adjacent operator,  $=\pi J_{\text{IS}} t 2 \hat{\mathbf{I}}_z \hat{\mathbf{S}}_z =>$ , in (2.4.2) yields the identity operator

$$=\omega_1 t \hat{\mathbf{I}}_z => =\pi \hat{\mathbf{I}}_\Phi => =\hat{\mathbf{I}}_E => =\omega_1 t \hat{\mathbf{I}}_z =>. \quad (2.4.4)$$

Removing the identity operator, we obtain

$$=\omega_1 t \hat{\mathbf{I}}_z => =\pi \hat{\mathbf{I}}_\Phi => =\omega_1 t \hat{\mathbf{I}}_z =>. \quad (2.4.5)$$

Placing the same identity operator,  $=\pi \hat{\mathbf{I}}_\Phi => =-\pi \hat{\mathbf{I}}_\Phi =>$ , at the beginning of sequence (2.4.5), one obtains

$$=\pi \hat{\mathbf{I}}_\Phi => \{-\pi \hat{\mathbf{I}}_\Phi => =\omega_1 t \hat{\mathbf{I}}_z => =\pi \hat{\mathbf{I}}_\Phi =>\} =\omega_1 t \hat{\mathbf{I}}_z =>. \quad (2.4.6)$$

Reducing the bracketed sequence yields

$$=\pi \hat{\mathbf{I}}_\Phi => =-\omega_1 t \hat{\mathbf{I}}_z => =\omega_1 t \hat{\mathbf{I}}_z =>. \quad (2.4.7)$$

Again the adjacent, opposite rotations constitute an identity operator and the sequence simplifies to

$$=\pi \hat{\mathbf{I}}_\Phi =>. \quad (2.4.8)$$

The evolution of a coupled two-spin system with this pulse sequence is identical to that of an isolated spin (Sequence 2.3.6).



### 2.5. Coupled two-spin system — simultaneous *I* and *S* 180° pulses

For a sequence with 180° pulses on both coupled spins (Table 2.1B), the appropriate rotation operators are

$$=\omega_I t \hat{I}_z => =\pi J_{IS} t 2 \hat{I}_z \hat{S}_z => =\pi (\hat{I}_\phi + \hat{S}_\phi) => =\pi J_{IS} t 2 \hat{I}_z \hat{S}_z => =\omega_I t \hat{I}_z =>. \quad (2.5.1)$$

The resulting composite rotation is composed of three simple rotations

$$=\pi \hat{S}_\phi => =\pi \hat{I}_\phi => =\pi J_{IS} 2 t 2 \hat{I}_z \hat{S}_z =>. \quad (2.5.2)$$

The 180° pulse on the *S* spin can be eliminated since it does not affect the detected *I* spin, giving

$$=\pi \hat{I}_\phi => =\pi J_{IS} 2 t 2 \hat{I}_z \hat{S}_z =>. \quad (2.5.3)$$

In addition to the phase shift arising from the 180° pulse, there is an evolution due to the active coupling interaction. The CFN descriptor for this evolution (Table 2.1B) is a flow primitive **E** that is combined with an active coupling modifier **A** connecting the spin tier of the transverse spin, *I*, with that of its coupling partner, *S*. Active coupling causes the evolution of inphase coherence into an antiphase state (or *vice versa*) between the coupled spins. For a spin with more than one coupling partner, multiple coupling modifiers connect the evolving coherence with the respective spin tiers.

### 2.6. Coupled two-spin system - selective *S* 180° pulse

The two other sequences (Table 2.1C and 2.1D) do not have a 180° pulse on the transverse component *I* and thus there is chemical shift evolution. The sequence in Table 2.1C represents a spin evolving under the chemical shift operator, Table 1.1CS, while the coupling interaction to *S* is refocused (decoupled) by the application of a 180° pulse to *S*. The rotations for this sequence are

$$=\omega_I t \hat{I}_z => =\pi J_{IS} t 2 \hat{I}_z \hat{S}_z => =\pi \hat{S}_\phi => =\pi J_{IS} t 2 \hat{I}_z \hat{S}_z => =\omega_I t \hat{I}_z =>. \quad (2.6.1)$$

After the insertion of the appropriate identities, the composite rotation arising from this sequence is

$$=\omega_I 2t \hat{\mathbf{I}}_z = >. \quad (2.6.2)$$

This is the same rotation obtained for the chemical shift evolution of an isolated spin. Notice that if there is a  $180^\circ$  pulse applied to either spin of the coupled pair of spins, the coupling interaction is eliminated.

## 2.7. Coupled two-spin system — no $180^\circ$ pulses

Table 2.1D represents a spin evolving under the chemical shift operator, Table 1.1CS, with active coupling, Table 1.1A, to another spin. The series of rotations for this sequence is

$$=\omega_I t \hat{\mathbf{I}}_z = > = \pi J_{IS} t 2 \hat{\mathbf{I}}_z \hat{\mathbf{S}}_z = >. \quad (2.7.1)$$

Both chemical shift and coupling are active during this sequence.

The use of the CFN description gives a visualization of the mechanism involved in multipulsed, multinuclear experiments. By introducing a graphical nomenclature, complicated experiments can be explained simply. In the remainder of this paper, the CFN diagrams will be used as an aid in the description of two-dimensional experiments. The utility of this description becomes even more apparent for three- and higher-dimensional experiments, in which evolutions on different nuclei occur simultaneously and the number of pulses and delays in the sequence is large.

## 3. PHASE CYCLING AND PULSED FIELD GRADIENTS

### 3.1. Introduction to phase cycling — quadrature images

The lack of a description of coherence selection in CFN diagrams is not to minimize the importance of phase cycling and pulsed field gradients (PFGs) in modern NMR spectroscopy. This apparent deficiency reflects the natural separation of coherence flow and coherence selection. The selection of coherence transfer pathways (CTP) by phase cycling or PFGs is the tool that allows the observation of coherence that has traversed desired pathway. This is separate from the generation of a specific coherence flow network. Once a pulse sequence is devised to generate a desired coherence, it remains to actively select the desired pathway through manipulation of the phase of the coherences; this is

the realm of phase cycling and PFG. Since this article focuses on the methods of 2D NMR experiments, which depend heavily on coherence selection, it is useful to describe, in detail, the procedure of phase cycling and PFGs for the removal of unwanted signals.

Coherence selection procedures play an integral role in 2D and ND experiments. The signals that are desired are usually “contaminated” by signals arising from unwanted coherences, imperfections in the RF pulses, errors in the phase of the receiver reference frequency, and other sources of artifacts. The selection of the desired coherence and the suppression of artifacts arising from instrumental problems is accomplished by using the dependence of coherence orders on the phase of RF pulses.

As an introduction to phase cycling, let us consider the removal of some imperfections caused by instrumental problems, specifically quadrature images. The nuclear spins generate an oscillating signal in the RF coil that can be described as

$$M(e^{i\Omega t} + e^{-i\Omega t}) = 2M \cos(\Omega t) \quad (3.1.1)$$

where  $M$  is the magnitude and  $\Omega$  is the radio frequency (RF). The signal is demodulated by mixing it with the carrier RF,  $\Omega_0$ . The process of demodulation subtracts the RF carrier signal from the received signal, producing the audio frequencies, which carry the information from the nuclear spins. The NMR detector is mathematically equivalent to  $e^{-i\Omega_0 t}$ , a circularly polarized RF signal [17]. Detection can be thought of as the product of the NMR signal with the detector:

$$M(e^{i\Omega t} + e^{-i\Omega t}) * e^{-i\Omega_0 t} \quad (3.1.2)$$

or

$$Me^{i(\Omega-\Omega_0)t} + Me^{-i(\Omega+\Omega_0)t} \quad (3.1.3)$$

In Equation (3.1.3),  $(\Omega+\Omega_0)$  is the sum of two RF signals which is another RF signal; this signal is unwanted and is filtered out electronically. The difference term,  $(\Omega-\Omega_0)$ , is the audio frequency that contains the information of the spin system.

The NMR detector is not ideal and pure circularly polarized radiation is not achievable. In the real world, the detector is contaminated with some amount of the oppositely rotating component,  $e^{i\Omega_0 t}$ . This is the

effect from the quadrature reference signals not having a relative phase shift of  $90^\circ$ . A real detector can be represented by:

$$e^{-i\Omega_0 t} + \lambda e^{i\Omega_0 t} \quad (3.1.4)$$

Using this detector, we obtain from the demodulation of the NMR signal:

$$M(e^{i\Omega t} + e^{-i\Omega t}) * (e^{-i\Omega_0 t} + \lambda e^{i\Omega_0 t}), \quad (3.1.5)$$

which can be expanded to:

$$Me^{i(\Omega-\Omega_0)t} + \lambda Me^{i(\Omega+\Omega_0)t} + Me^{-i(\Omega+\Omega_0)t} + \lambda Me^{-i(\Omega-\Omega_0)t}. \quad (3.1.6)$$

As before the terms containing  $\Omega+\Omega_0$  are RF frequencies that are eliminated by filters, but the audio frequency component,  $\Omega-\Omega_0$ , contains two terms at  $\pm(\Omega-\Omega_0)$ . Fourier transform of these signals will produce two peaks: the correct peak with magnitude  $M$  will appear at  $(-\Omega-\Omega_0)$ , but another peak with amplitude  $\lambda M$  will appear at the symmetric position about the carrier frequency,  $(\Omega-\Omega_0)$ . This false signal is a quadrature image and is clearly undesirable. A phase cycle known as CYCLOPS is used to suppress these artifacts. The sequence consists of collecting and combining in a prescribed manner, the signals from 4 individual experiments. The phases of the excitation pulse and receiver are incremented by  $90^\circ$  in each experiment.

$$\begin{array}{l}
 90(0) \\
 \text{_____} \blacksquare \rangle \rangle \rangle \text{ (receiver } 0) \\
 \\
 90(\pi/2) \\
 \text{_____} \blacksquare \rangle \rangle \rangle \text{ (receiver } 3\pi/2) \\
 \\
 90(\pi) \\
 \text{_____} \blacksquare \rangle \rangle \rangle \text{ (receiver } \pi) \\
 \\
 90(3\pi/2) \\
 \text{_____} \blacksquare \rangle \rangle \rangle \text{ (receiver } \pi/2)
 \end{array} \quad (3.1.7)$$

The sequences in (3.1.7) are a schematic of the CYCLOPS phase cycle procedure. In the following sections, we will analyze this procedure in detail, but we first must discuss an implementation of the product operators that is convenient for the discussion of phase cycling.

### 3.2. Raising and lowering operators

In order to understand phase cycling, it is useful to use an alternate description of the product operators. The raising,  $\mathbf{I}^+$ , and lowering,  $\mathbf{I}^-$ , operators are well known in spectroscopy. Whereas the  $\mathbf{I}_x$  and  $\mathbf{I}_y$  operators involve both up and down transitions between energy levels,  $\mathbf{I}^+$  and  $\mathbf{I}^-$  only allow transitions in one direction.  $\mathbf{I}^+$  is the transition up and  $\mathbf{I}^-$  is the transition down.

These operators are defined as:

$$\begin{aligned}\mathbf{I}^+ &\equiv (\mathbf{I}_x + i\mathbf{I}_y) \\ \mathbf{I}^- &\equiv (\mathbf{I}_x - i\mathbf{I}_y).\end{aligned}\tag{3.2.1}$$

Combining and rearranging Equations (3.2.1), one obtains

$$\begin{aligned}\mathbf{I}_x &= 1/2(\mathbf{I}^+ + \mathbf{I}^-) \\ \mathbf{I}_y &= 1/(2i)(\mathbf{I}^+ - \mathbf{I}^-) = -i/2(\mathbf{I}^+ - \mathbf{I}^-)\end{aligned}\tag{3.2.2}$$

These operators have useful properties for the description of the evolution of a spin system subjected to phase shifts. For example, consider the effect of a phase shift,  $=\phi\mathbf{I}_z=>$ , on the  $\mathbf{I}^+$  and  $\mathbf{I}^-$  operators,

$$\mathbf{I}^+ = \phi \hat{\mathbf{I}}_z => ?\tag{3.2.3}$$

This can easily be calculated by using the known rotations of the Cartesian product operators

$$(\mathbf{I}_x + i\mathbf{I}_y) = \phi \hat{\mathbf{I}}_z => \mathbf{I}_x \cos\phi + \mathbf{I}_y \sin\phi + i\mathbf{I}_y \cos\phi - i\mathbf{I}_x \sin\phi.\tag{3.2.4}$$

This can be simplified to

$$\mathbf{I}_x (\cos\phi - i\sin\phi) + i\mathbf{I}_y (\cos\phi - i\sin\phi) = \mathbf{I}_x e^{-i\phi} + i\mathbf{I}_y e^{-i\phi}\tag{3.2.5}$$

or further to

$$(\mathbf{I}_x + i\mathbf{I}_y) e^{-i\phi} = \mathbf{I}^+ e^{-i\phi}. \quad (3.2.6)$$

In summary, Eq. (3.2.3) becomes

$$\mathbf{I}^+ = \phi \hat{\mathbf{I}}_z \Rightarrow \mathbf{I}^+ e^{-i\phi}. \quad (3.2.7)$$

Likewise, the transformation for the  $\mathbf{I}^-$  operator is

$$\mathbf{I}^- = \phi \hat{\mathbf{I}}_z \Rightarrow \mathbf{I}^- e^{i\phi}. \quad (3.2.8)$$

A phase shift,  $=\phi \hat{\mathbf{I}}_z \Rightarrow$ , is equivalent to the chemical shift operator,  $=\Omega_1 t \hat{\mathbf{I}}_z \Rightarrow$ , and therefore, the evolution of  $\mathbf{I}^+$  and  $\mathbf{I}^-$  under chemical shift can be described as:

$$\mathbf{I}^+ = \Omega_1 t \hat{\mathbf{I}}_z \Rightarrow \mathbf{I}^+ e^{-i\Omega t}$$

and

$$\mathbf{I}^- = \Omega_1 t \hat{\mathbf{I}}_z \Rightarrow \mathbf{I}^- e^{i\Omega t}.$$

The useful property of the raising and lowering operators is that they do not mix with a Z rotation. It is the behavior of these operators under Z rotations (phase shift and chemical shift) that makes them convenient for describing the behavior of spins in a phase cycled experiment.

### 3.3. Phase shifted RF pulses

Phase shifted RF pulses can be thought of as a composite pulse of an X axis pulse followed by a Z axis “pulse”.

$$\mathbf{I}_z = \pi/2 \hat{\mathbf{I}}_x \Rightarrow -\mathbf{I}_y \quad (3.3.1)$$

This is the same as a  $90^\circ_x$  followed by a  $0^\circ_z$  pulse,

$$\mathbf{I}_z = \pi/2 \hat{\mathbf{I}}_x \Rightarrow -\mathbf{I}_y = 0 \hat{\mathbf{I}}_z \Rightarrow -\mathbf{I}_y = i/2(\mathbf{I}^+ - \mathbf{I}^-). \quad (3.3.2)$$

The equivalent state using the raising and lowering operators is given to show the effects of the phase shift.

With a  $\pi/2$  phase shift of a  $\hat{\mathbf{I}}_x$  pulse, we obtain a  $\hat{\mathbf{I}}_y$  pulse. This can be simulated as an X pulse followed by a  $90^\circ$  Z rotation,

$$\mathbf{I}_z = \pi/2 \hat{\mathbf{I}}_y \Rightarrow \mathbf{I}_x = \mathbf{I}_z = \pi/2 \hat{\mathbf{I}}_x \Rightarrow \pi/2 \hat{\mathbf{I}}_z \Rightarrow \mathbf{I}_x = 1/2(\mathbf{I}^+ + \mathbf{I}^-). \quad (3.3.3)$$

For a  $90^\circ$  phase shift, the  $\mathbf{I}^+$  component is multiplied by  $-i$  and the  $\mathbf{I}^-$  component is multiplied by  $i$ .

The  $180^\circ$  (Eq. 3.3.4) and  $270^\circ$  (Eq. 3.3.5) phase shifts can be described similarly:

$$\mathbf{I}_z = \pi/2 \hat{\mathbf{I}}_x \Rightarrow \mathbf{I}_y = \mathbf{I}_z = \pi/2 \hat{\mathbf{I}}_x \Rightarrow \pi \hat{\mathbf{I}}_z \Rightarrow \mathbf{I}_y = -i/2(\mathbf{I}^+ - \mathbf{I}^-) \quad (3.3.4)$$

$$\mathbf{I}_z = \pi/2 \hat{\mathbf{I}}_y \Rightarrow \mathbf{I}_x = \mathbf{I}_z = \pi/2 \hat{\mathbf{I}}_x \Rightarrow 3\pi/2 \hat{\mathbf{I}}_z \Rightarrow \mathbf{I}_x = -1/2(\mathbf{I}^+ + \mathbf{I}^-). \quad (3.3.5)$$

### 3.4. Elimination of quadrature images

As discussed in Section 3.1, if the signal is detected by a contaminated carrier signal then we obtain quadrature images. The signal for a X axis pulse given by

$$\lambda \text{Me}^{-i(\Omega - \Omega_0)t} + \text{Me}^{i(\Omega - \Omega_0)t} \quad (3.4.1)$$

Based on the phase shift behavior, we equate the NMR signal Eq. (3.1.1),  $\text{Me}^{i\Omega t}$  with  $M \mathbf{I}^-$  and  $e^{-i\Omega t}$  with  $M \mathbf{I}^+$ . With a  $90^\circ$  phase shift in the RF pulse, the  $\mathbf{I}^+$  component or its equivalent ( $e^{-i\Omega - \Omega_0 \tau}$ ) is multiplied by  $-i$  and the  $\mathbf{I}^-$  component,  $e^{i(\Omega - \Omega_0)\tau}$ , is multiplied by  $i$  (Eqn 3.3.3). Multiplication of the appropriate terms in Eq. (3.4.1) with  $-i$  and  $i$  results in

$$-i\lambda \text{Me}^{-i(\Omega - \Omega_0)t} + i\text{Me}^{i(\Omega - \Omega_0)t} \quad (3.4.2)$$

With a  $180^\circ$  phase shift in the RF we obtain

$$-\lambda \text{Me}^{-i(\Omega - \Omega_0)t} - \text{Me}^{i(\Omega - \Omega_0)t} \quad (3.4.3)$$

and for a  $270^\circ$  phase shift:

$$i\lambda \text{Me}^{-i(\Omega - \Omega_0)t} - i\text{Me}^{i(\Omega - \Omega_0)t} \quad (3.4.4)$$

The signals are acquired separately and combined in computer memory. However, we can not directly add these results since half are real and the other half are imaginary. Equations (3.4.1) and (3.4.3) and Eqs. (3.4.2) and (3.4.4) are negatives of one another and need to be subtracted to retain the desired signal  $\text{Me}^{i(\Omega-\Omega_0)t}$ . In order to add Eqs. (3.4.1) and (3.4.3) to Eqs. (3.4.2) and (3.4.4), we need to multiply the even numbered equations by  $i$  to convert them to pure real. This operation is equivalent to phase shifting the receiver. Recalling Euler's formula of

$$e^{\pm iA} = \cos A \pm i \sin A \quad (3.4.5)$$

we obtain

$$\begin{aligned} e^{i0} &= 1, e^{i3\pi/2} = -i, \\ e^{i\pi} &= -1, \text{ and } e^{i\pi/2} = i, \end{aligned} \quad (3.4.6)$$

which correspond to shifting the receiver phase by 0,  $3\pi/2$ ,  $\pi$ , and  $\pi/2$ , multiplication of Eqs. (3.4.1) and (3.4.3) by 1 and  $-1$ , and multiplication of Eqs. (3.4.2) and (3.4.4) by  $-i$  and  $i$ , respectively, we obtain

$$\begin{aligned} &\lambda \text{Me}^{-i(\Omega-\Omega_0)t} + \text{Me}^{i(\Omega-\Omega_0)t} \\ &- \lambda \text{Me}^{-i(\Omega-\Omega_0)t} + \text{Me}^{i(\Omega-\Omega_0)t} \\ &\lambda \text{Me}^{-i(\Omega-\Omega_0)t} + \text{Me}^{i(\Omega-\Omega_0)t} \\ &- \lambda \text{Me}^{-i(\Omega-\Omega_0)t} + \text{Me}^{i(\Omega-\Omega_0)t}. \end{aligned} \quad (3.4.7)$$

Adding Eqs. (3.4.7), we are left with

$$4 * \text{Me}^{i(\Omega-\Omega_0)t} \quad (3.4.8)$$

The contaminating component,  $\lambda \text{Me}^{-i(\Omega-\Omega_0)t}$ , is eliminated along with the associated quadrature image in the spectrum. The desired component  $\text{Me}^{i(\Omega-\Omega_0)t}$  is retained and has sum of the amplitudes from all experiments.

### 3.5. Phase shifts by data routing

The actual “phase shifting” of the receiver usually is not accomplished by RF phase shifts, but by specific data routing in the computer memory. As an example, consider the signal:

$$e^{i\omega t} = \cos \omega t + i * \sin \omega t \quad (3.5.1)$$



where the cosine and sine functions represent the digitized output of two orthogonal receiver channels X (real) and Y (imaginary), respectively. A  $90^\circ$  phase shift is represented as  $e^{i\pi/2} = i$ . Multiplying the signal (3.5.1) with  $i$  yields

$$i * e^{i\omega t} = i * (\cos \omega t + i * \sin \omega t) = i * \cos \omega t - \sin \omega t. \quad (3.5.2)$$

The X channel, represented by the cosine function, is now the imaginary part of the signal and the Y channel, represented by the sine function, is the negative of the real part of the signal. A  $90^\circ$  phase shift in the receiver is equivalent to swapping the real and imaginary channels combined with the negation of one of them. For CYCLOPS the X and Y channels are routed to two separate memory locations in the computer representing the real and imaginary parts of the complex pair as follows:

$\phi$	A	B
$0^\circ$	X	Y
$90^\circ$	-Y	X
$180^\circ$	-X	-Y
$270^\circ$	Y	-X

where the negative signs mean that the signal is subtracted from memory. This method is a perfect  $90^\circ$  phase shift and cannot introduce errors that would occur if implemented in hardware. By taking other linear combinations of the X and Y channels, arbitrary "receiver" phase shifts can be generated in computer memory.

### 3.6. The rules of phase cycling [22]

Specific rules are available for the construction of phase cycling regimens. By following these rules, the phase cycle for the selection of a given coherence transfer pathway can be constructed for any pulse sequence. In summary, the rules are:

- (1) Write down all possible changes in coherence,  $\Delta m$ , due to the pulse.
- (2) Mark the desired coherence change and place a bracket next to it.
- (3) Place a closing bracket after the last undesired coherence change.
- (4) Count the number of terms N inside the brackets.
- (5) Calculate the proper phase shifts using Eq. (3.6.1).

$$\phi_i = 2\pi k/N \quad k = 0, 1, \dots, N-1 \quad (3.6.1)$$

- (6) Repeat steps (1) through (5) for all other pulses in the sequence.  
 (7) The receiver phase,  $\psi$ , is calculated for each phase cycle step  $i$  by Eq. (3.6.2).

$$\psi = -\sum \Delta m_i \phi_i \quad (3.6.2)$$

where  $\Delta m_i$  is the coherence order change at the  $i^{\text{th}}$  pulse and  $\phi_i$  is the phase of that pulse.

Several common phase cycles will now be developed using these rules.

### 3.7. Automatic baseline compensation or systematic noise reduction

A common problem is data that is biased by a constant signal. This type of signal can arise from DC offset in amplifiers or other systematic noise. The elimination of this type of artifact relies on the constant nature of the signal.

Systematic noise reduction:



The diagram in (3.7.1) has the pulse sequence on top and below is the coherence transfer pathway (CTP) [22]. The double lines in (3.7.1) mark the desired coherence transfer pathway. Single lines are coherence orders that may occur but are undesired. The x marks the pathway that is to be blocked. Magnetization at thermal equilibrium has coherence order 0 (level 0). Pulses cause transitions between levels while chemical shift and coupling evolution do not change the coherence order. By convention, we detect coherence order  $-1$  as indicated by the box at level  $-1$ . With a perfect detector, only  $-1$  coherence is observed.

Following the procedure of Section 3.6:

- (1) Possible changes in coherence due to the pulse

$$\Delta m = +1$$

$$\Delta m = 0$$

$$\Delta m = -1.$$

At thermal equilibrium the coherence order is 0. Only  $\pm 1$  coherence can be excited from thermal equilibrium.

- (2) Mark the desired coherence change and place a bracket next it. Here the desired change is set in boldface.

$$+1, 0, -1)$$

- (3) Place a closing bracket after last undesired coherence change, here we want to suppress 0 order coherence. A change of +1 can be ignored since with a perfect detector only the  $-1$  coherence is detected.

$$+1,(0, -1)$$

- (4) Count the number of terms, N, inside the brackets:

$$N=2$$

- (5) Calculate RF phase shifts for this pulse using Eq. (3.6.1).

$$\phi_k = 2\pi k/N \quad k = 0, 1, \dots, N-1$$

For  $N=2$ ,  $k$  has values of 0 and 1.

$$\phi_1 = 2\pi * 0/2 = 0$$

$$\phi_2 = 2\pi * 1/2 = \pi$$

- (6) The receiver phase is calculated from Eq. (3.6.2).

$$\psi = -\sum \Delta m_i \phi_i$$

In this case,

$$\psi_1 = -(-1)*0$$

$$\psi_2 = -(-1)*\pi$$

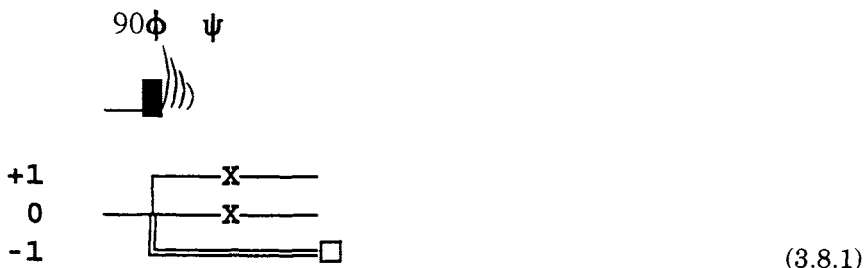
The complete phase cycling is given by the following table:

$\phi$ :	0	$\pi$
$\psi$ :	0	$\pi$

Two experiments are collected. The phase of the pulse  $\phi$  is changed from 0 to  $\pi$  with a corresponding change in the receiver phase of 0 to  $\pi$ . The phase shift in the receiver is accomplished simply by subtracting the signals arising from the two pulses.

### 3.8. Elimination of quadrature images and DC offset

We now return to the elimination of quadrature images and the development of the CYCLOPS procedure (Section 3.4).



In sequence (3.8.1), a single  $90^\circ$  pulse excites Z magnetization at coherence order 0 to orders  $\pm 1$ . To eliminate quadrature images  $+1$  coherence must be blocked. Removal of systematic noise is accomplished by blocking 0 order coherence.

Following the procedure of Section 3.6:

- (1) List the possible coherence changes.

+1, 0, -1

- (2) Mark the desired coherence change and place a bracket next to it.

+1, 0, -1)

- (3) Place a closing bracket after last undesired coherence change.

(+1, 0, -1)

Note: +1 must be eliminated with an imperfect detector.

(4) Count the number of terms inside the brackets.

$$N=3.$$

(5) The RF phase shifts for this pulse are

$$\phi_1 = 2\pi \cdot 0/3 = 0$$

$$\phi_2 = 2\pi \cdot 1/3 = 2\pi/3$$

$$\phi_3 = 2\pi \cdot 2/3 = 4\pi/3.$$

(6) Calculate the receiver phases,

$$\psi_1 = -(-1) \cdot 0 = 0$$

$$\psi_2 = -(-1) \cdot 2\pi/3 = 2\pi/3$$

$$\psi_3 = -(-1) \cdot 4\pi/3 = 4\pi/3$$

The phase cycle for the CTP in sequence (3.8.1) is

$$\phi: 0 \ 2\pi/3 \ 4\pi/3$$

$$\psi: 0 \ 2\pi/3 \ 4\pi/3.$$

This 3-step sequence is not the normal 4-step CYCLOPS sequence (Section 3.4), which is

$$\phi: 0 \ \pi/2 \ \pi \ 3\pi/2$$

$$\psi: 0 \ \pi/2 \ \pi \ 3\pi/2.$$

The main reason that the 4-step phase cycling routine arises is that  $90^\circ$  phase shifts are much easier to implement in hardware than are sub- $90^\circ$  phase shifts. The CYCLOPS procedure was invented long before sub- $90^\circ$  phase shifts were commonly available on spectrometers.

### 3.9. CYCLOPS [18]

To arrive at the normal CYCLOPS phase cycle, consider a fictitious level at coherence order +2 that is to be eliminated. This level of coherence cannot be generated from equilibrium magnetization and therefore, need not be eliminated. However, there is nothing preventing us from removing nonexistent things. The CTP of (3.7.1) is modified by the addition of a +2 coherence order level producing the CTP in (3.9.1).



- (1) List the possible coherence changes

+2, +1, 0, -1.

- (2) Mark the desired coherence change

+2, +1, 0, -1).

- (3) Place a closing bracket after last undesired coherence change

(+2, +1, 0, -1).

- (4) Count the number of terms inside the brackets

N=4.

- (5) The RF phase shifts for this pulse are

$$\phi_1 = 2\pi \cdot 0/4 = 0$$

$$\phi_2 = 2\pi \cdot 1/4 = \pi/2$$

$$\phi_3 = 2\pi \cdot 2/4 = \pi$$

$$\phi_4 = 2\pi \cdot 3/4 = 3\pi/2.$$

- (6) The receiver phases are

$$\psi_1 = -(-1) \cdot 0 = 0$$

$$\psi_2 = -(-1) \cdot \pi/2 = \pi/2$$

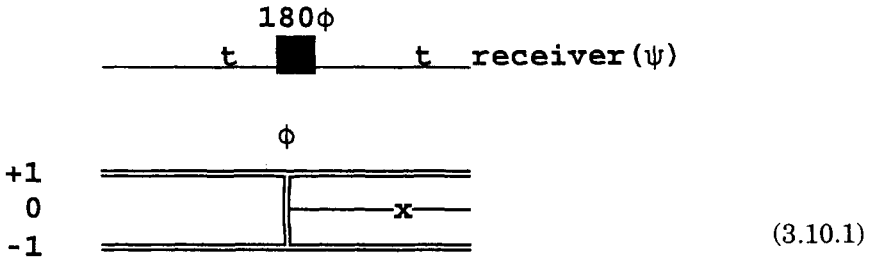
$$\psi_3 = -(-1) \cdot \pi = \pi$$

$$\psi_4 = -(-1) \cdot 3\pi/2 = 3\pi/2.$$

This is the normal CYCLOPS routine that is commonly used in spectrometers to reduce quadrature images. As discussed in Section 3.5, the receiver phase shifts are normally implemented through data routing.

### 3.10. EXORCYCLE [19]

EXORCYCLE is a phase cycling routine designed to remove imperfections in a  $180^\circ$  rotation of transverse magnetization. A  $180^\circ$  pulse on transverse coherence is equivalent to interchange of  $\mathbf{I}^*$  to  $\mathbf{I}$ . Coherence transfers of 0 and  $\pm 1$  need to be eliminated.



- (1) The desired coherence changes at the  $180^\circ$  pulse are  $\Delta m = \pm 2$ .
- (2) Writing all possible changes and labeling as above  $(+2 +1 0 -1) -2$ .
- (3) This gives  $N = 4$  terms inside the bracket, leading to phase shifts for  $\phi$  of  $\phi_1 = 0, \pi/2, \pi, 3\pi/2$ .
- (4) The receiver phase cycle is now calculated,

$$\begin{aligned}\psi_1 &= -(2) * 0 = 0 \\ \psi_2 &= -(2) * \pi/2 = \pi \\ \psi_1 &= -(2) * \pi = 2\pi = 0 \\ \psi_2 &= -(2) * 3\pi/2 = 3\pi = \pi.\end{aligned}$$

The resulting EXORCYCLE phase cycle is

$$\begin{aligned}\phi: & 0 \ \pi/2 \ \pi \ 3\pi/2 \\ \psi: & 0 \ \pi \ 0 \ \pi.\end{aligned}$$

To make the operation of this phase cycle a bit more transparent, let us use the product operators to show how this sequence works. Transverse magnetization is represented as:

$$-I_y \cos \omega_t t + I_x \sin \omega_t t \quad (3.10.2)$$

If the  $180^\circ$  pulse is misset, i.e., the actual pulse angle  $< 180^\circ$  by some

small angle  $\beta$ , here are the calculations for the four different RF phases on transverse magnetization.

$$=(\pi-\beta)\hat{\mathbf{I}}_{\mathbf{x}}=> (-\mathbf{I}_y \cos (\pi-\beta) - \mathbf{I}_z \sin (\pi-\beta)) \cos \omega_1 t + \mathbf{I}_x \sin \omega_1 t \quad (3.10.3)$$

$$=(\pi-\beta)\hat{\mathbf{I}}_{\mathbf{y}}=> -\mathbf{I}_y \cos \omega_1 t + (\mathbf{I}_x \cos (\pi-\beta) - \mathbf{I}_z \sin (\pi-\beta)) \sin \omega_1 t \quad (3.10.4)$$

$$=(\pi-\beta)\hat{\mathbf{I}}_{\mathbf{x}}=> (-\mathbf{I}_y \cos (\pi-\beta) + \mathbf{I}_z \sin (\pi-\beta)) \cos \omega_1 t + \mathbf{I}_x \sin \omega_1 t \quad (3.10.5)$$

$$=(\pi-\beta)\hat{\mathbf{I}}_{\mathbf{y}}=> -\mathbf{I}_y \cos \omega_1 t + (\mathbf{I}_x \cos (\pi-\beta) + \mathbf{I}_z \sin (\pi-\beta)) \sin \omega_1 t \quad (3.10.6)$$

In sequences (3.10.3) and (3.10.5), and in sequences (3.10.4) and (3.10.6) the  $\mathbf{I}_z$  terms are opposite in sign. Adding the acquired data from these pairs of sequences will eliminate the  $\mathbf{I}_z$  component. These pairs are then subtracted to retain the desired transverse operators,  $\mathbf{I}_x$  and  $\mathbf{I}_y$ . The elimination of the  $\mathbf{I}_z$  term is not important if this sequence is followed by the acquisition period, since Z magnetization is not detected. If this sequence, however, is an element of a longer pulse sequence, other pulses can follow the spin echo segment. These pulses will rotate the undesired  $\mathbf{I}_z$  into the transverse plane and cause artifacts to appear in the spectrum. This exercise shows that in any individual step of a phase cycle unwanted terms are retained, and only through combinations of phase shifted experiments, consisting of RF phase shifts and addition or subtraction in computer memory, do the desired terms co-add and the undesired terms cancel.

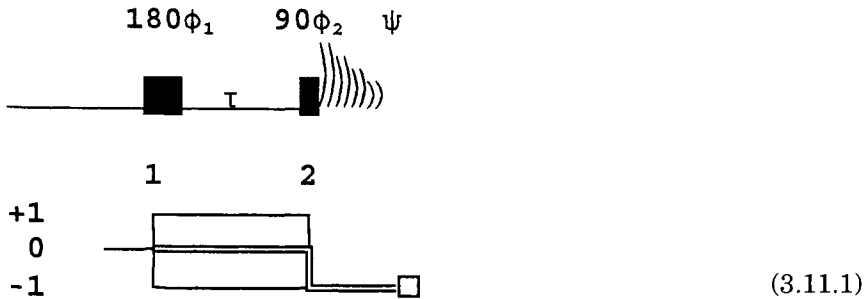
Another scenario that can give rise to artifacts arises if, during the first evolution period in the spin echo sequence, some Z magnetization is generated from longitudinal relaxation. An imperfect  $180^\circ$  pulse then generates unwanted transverse coherence from the Z magnetization. The EXORCYCLE sequence eliminates  $\pm 1$  coherence transfer; and, as can be shown with product operator analysis, the interfering transverse component is eliminated.

Phase cycling does not eliminate the undesired coherences in a single acquisition; but is only effective upon addition of the separate experiments in the computer memory. However, with respect to the interpretation of the pulse sequence, the EXORCYCLE phase cycle can be considered to eliminate the  $\mathbf{I}_z$  component at this point in the sequence. *No matter what occurs in the pulse sequence beyond this point, any signals that arise from the  $\mathbf{I}_z$  terms will remain opposite in sign and be canceled in the computer memory.*



### 3.11. Inversion of longitudinal magnetization

A major use for inversion of Z magnetization is to decouple (or recouple) the scalar coupling interaction between two spins (Section 2.6). If the inversion is not perfect, transverse magnetization is produced which can interfere with the desired coherences. The use of longitudinal inversion pulses is also important in the measurement of relaxation parameters. In the inversion recovery sequence, Sequence (3.11.1), longitudinal magnetization +Z is ideally inverted to pure -Z magnetization. Imperfections in the pulse may cause some transverse magnetization to appear. Phase cycling is used to eliminate this unwanted coherence. The desired change in coherence order is 0 in this case since both +Z and -Z are of order 0.



Proceeding as before:

$$\begin{aligned} \phi_1: \\ \Delta m &= 0 \\ (+1 \ 0) - 1 \\ N &= 2 \end{aligned}$$

The phases for the inversion  $180^\circ$  pulse are found.

$$\begin{aligned} \phi_1 &= 2\pi k/2 \ (k=0,1) \\ \phi_1 &= 0, 2 \end{aligned}$$

The receiver phase for all RF phases is then

$$\psi_1 = -(0) * \phi_1 = 0.$$

If the experiment is repeated too rapidly such that there are transverse coherences that have not completely relaxed to equilibrium between experiments, then we will have coherence order changes of  $\pm 2$ . If  $\Delta m = +2$  is also to be eliminated then

$$\phi_1: \begin{pmatrix} +2 & +1 & 0 & -1 \end{pmatrix}$$

$$N=4$$

and the rf pulse phases are

$$\phi_1 = 0, \pi/2, \pi, 3\pi/2$$

with all of the receiver phases being

$$\psi_1 = -(0) * \phi_1 = 0.$$

The  $90^\circ$  read pulse transfers the Z magnetization to the transverse plane for detection and so the desired coherence order change is  $-1$ . This pulse can be phased cycled with the CYCLOPS sequence to eliminate quadrature images (Section 3.9).

The phase cycles for complicated experiments with many pulses and many different orders of coherences are easily derived by following the above rules. The procedure in Section 3.6 will be used to develop phase cycling regimens for the two-dimensional experiments described in this article.

### ***3.12. Pulsed field gradients and coherence selection***

Recently, an old technique has been introduced as a complement to phase cycling [20]. Pulsed field gradients can be used to select (or deselect) coherence transfer pathways. Recent advances in the use of pulsed field gradients for multidimensional spectroscopy have spurred the manufacturers to supply high resolution probes fitted with gradient coils.

The principle on which this technique is based is the phase sensitivity of different coherence orders. A coherence when subjected to a phase shift has a resultant phase dependent upon the order of coherence. For example,

Coherence order 1:

$$\mathbf{I}^+ = \phi \hat{\mathbf{I}}_z \Rightarrow \mathbf{I}^+ e^{-i\phi} \quad (3.12.1)$$

Coherence order 2:

$$\mathbf{I}^+\mathbf{S}^+ = \phi \hat{\mathbf{I}}_z + \phi \hat{\mathbf{S}}_z \Rightarrow \mathbf{I}^+\mathbf{S}^+ e^{-i2\phi} \quad (3.12.2)$$

Coherence order 0:

$$\mathbf{I}^+\mathbf{S}^+ = \phi \hat{\mathbf{I}}_z + \phi \hat{\mathbf{S}}_z \Rightarrow \mathbf{I}^+\mathbf{S}^+ e^0 \quad (3.12.3)$$

$$\mathbf{I}_z = \phi \hat{\mathbf{I}}_z \Rightarrow \mathbf{I}_z e^0 \quad (3.12.4)$$

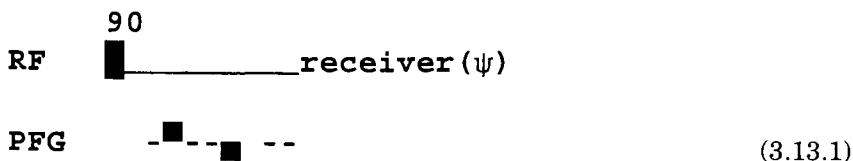
In general, a coherence with order  $\pm p$  will experience a phase shift according to

$$\mathbf{I}_1^\pm \mathbf{I}_2^\pm \dots \mathbf{I}_p^\pm = \phi \hat{\mathbf{I}}_z \Rightarrow (\mathbf{I}_1^\pm \mathbf{I}_2^\pm \dots \mathbf{I}_p^\pm) e^{\mp i p \phi} \quad (3.12.5)$$

This sensitivity to phase shift is the property that is used in phase cycling. In phase cycling, however, the phase of the RF pulses is shifted to effect a phase shift in the coherence. The phase shift of the desired coherence is then followed by an appropriate shift in the receiver phase such that the signal arising from the desired coherence constructively interferes with signals from previous experiments. If a radiofrequency Z pulse were accessible, then the coherence itself could be phase shifted. A Z pulse could be accomplished by a transient change in the Larmor frequency. This could be implemented experimentally by changing the magnetic field strength for a specified time. A spin would experience a higher (or lower) magnetic field strength and would precess at a different frequency. If calibrated properly, an arbitrary phase shift, governed by Eq. (3.12.5), could be obtained. A more practical use of changes in the magnetic field strength is to use a magnetic field gradient. If a spatially inhomogeneous magnetic field is applied to a sample, then the same spin on different molecules in different parts of the sample will feel different magnetic field strengths and precess at different frequencies.

If a  $90^\circ$  pulse is applied to a spin system and then a pulsed magnetic field gradient (PFG) is applied, transverse coherences in different magnetic fields (different parts of the sample) are subjected to a phase shift dependent on the strength and duration of the PFG. After the PFG the transverse spin vectors throughout the sample are dephased and, since the detected signal comes from all parts of the sample, the integrated signal is zero. The dephasing, however, can be reversed by several methods, to be described below.

### 3.13. Gradient recalled echoes



By reversing the direction of a dephasing gradient, the transverse coherences can be rephased and the signal recovered in a gradient recalled echo. For a transverse coherence  $\mathbf{I}^+$  generated by the  $90^\circ$  pulse in Sequence (3.13.1), the first PFG dephases the coherence:

$$\mathbf{I}^+ = \gamma G(r) t \hat{\mathbf{I}}_z \Rightarrow \mathbf{I}^+ e^{-i\gamma G(r)t} \quad (3.13.2)$$

where  $\gamma$  is the magnetogyric ratio of the nucleus and  $G(r)$  represents the magnitude and distribution of the gradient field. By applying an opposite and equal PFG the coherence is recovered:

$$\mathbf{I}^+ e^{-i\gamma G(r)t} = -\gamma G(r) t \hat{\mathbf{I}}_z \Rightarrow \mathbf{I}^+ e^{-i\gamma G(r)t} * e^{i\gamma G(r)t} = \mathbf{I}^+ \quad (3.13.3)$$

If the time between the application of the PFGs is long enough for something to happen to the spin such that it does not “see” a field of the same magnitude and opposite polarity, then the coherence remains dephased and does not contribute to the final detected signal. For example, if molecular diffusion occurs such that during the time between the PFGs the molecule moves from its initial position in the sample to a different position that experiences a significantly different magnetic field during the second PFG, then the rephasing gradient will not be completely effective in recovering the signal (Eq. 3.13.4).

$$\mathbf{I}^+ = \gamma G(r) t \hat{\mathbf{I}}_z \Rightarrow \mathbf{I}^+ e^{-i\gamma G(r)t} = -\gamma G'(r) t \hat{\mathbf{I}}_z \Rightarrow \mathbf{I}^+ e^{-i\gamma G(r)t} * e^{i\gamma G'(r)t} = \mathbf{I}^+ e^{i\gamma [G'(r) - G(r)]t} \quad (3.13.4)$$

The phase introduced by the first PFG is not canceled by the second PFG and thus the signal is not completely refocused. This effect has been used for solvent suppression, taking advantage of the rapid diffusion of water compared to macromolecules.



and finally the second PFG, which is identical to the first PFG, rephases the  $\mathbf{I}^-$  coherence.

$$\mathbf{I}^- e^{-i\gamma G(r)t} = -\gamma G(r)t \hat{\mathbf{I}}_z \Rightarrow \mathbf{I}^- e^{-i\gamma G(r)t} * e^{i\gamma G(r)t} = \mathbf{I}^- \quad (3.14.5)$$

The inverse sensitivity of the  $\mathbf{I}^+$  and  $\mathbf{I}^-$  coherences to the gradient induced phase shift causes the signal to refocus in the two identical PFGs. Any imperfection in the  $180^\circ$  pulse will produce  $\mathbf{I}_z$  magnetization from gradient-dephased transverse magnetization present before the RF pulse (Sequence 3.14.3). The undesired  $\mathbf{I}_z$  magnetization is labeled with the dephased transverse magnetization that occurred during the PFG.

$$\mathbf{I}^+ e^{-i\gamma G(r)t} = \beta \hat{\mathbf{I}}_x \Rightarrow \mathbf{I}_z e^{-i\gamma G(r)t} \sin(\beta) \quad (3.14.6)$$

If there is another RF pulse later in the experiment, the transverse magnetization created from the Z magnetization of Eq. (3.14.6) will retain the dephased information and, in the absence of a rephasing gradient, will not contribute to the final signal.

$$\mathbf{I}_z e^{-i\gamma G(r)t} = \pi/2 \hat{\mathbf{I}}_x \Rightarrow -\mathbf{I}_y e^{-i\gamma G(r)t} \quad (3.14.7)$$

For Z magnetization that is present before the imperfect  $180^\circ$  pulse, any generated transverse magnetization will be dephased by the second PFG.

Pulsed field gradients also can be used to remove the phase cycling necessary to select a given coherence. They can also be used to select either the  $+\mathbf{I}$  or  $-\mathbf{I}$  coherence for magnitude-mode quadrature detection. PFGs promise to be an extremely useful tool in NMR spectroscopy.

## 4. TWO-DIMENSIONAL NMR

### 4.1. Exchange spectroscopy (EXSY) [21]

The basic structure of all two-dimensional spectra can be divided into a sequence of four periods [1]. The *preparation period* is used to place the spin system into the desired starting conditions. The starting condition may range from a relaxed spin system at Boltzmann equilibrium to a multiple quantum state or spin order. This period is followed by the *evolution period*,  $t_1$ , during which the spins evolve at their

characteristic frequencies. This is the period that is incremented to generate the indirectly detected time domain. Since the  $t_1$  time domain is indirectly detected, the restrictions on the “observability” of the coherences that evolve during this time are lifted. The  $t_1$  period can witness the evolution of single quantum coherence as well as multiple quantum coherence, which is not directly observable. The next period in a two-dimensional NMR experiment is the *mixing period*. The transfer of information between spins is carried out during this period. This construction of this period is very diverse. It can range from being absent to containing very complicated sequences of RF pulses and delays. The mixing period is the link between the evolution period and the *detection period*. The spectrometer’s receiver is turned on during the detection period and the FID is detected and stored in the computer memory.

As a paradigm of a 2D experiment, I will describe a somewhat contrived system for the study of chemical exchange. Chemical exchange is especially convenient for the explanation of the basics of two-dimensional spectroscopy because the mechanism of (incoherent) magnetization transfer involves the simple physical movement of a nucleus from one chemical site to another. Since the transfer mechanism is classical in nature, the experiment can be described in the framework of classical physics. This simple experiment will be used to exemplify many general features of the 2D NMR experiment.

Consider a rapid, irreversible reaction,  $I \rightarrow S$ , which is induced by light. The protons on molecule **I** are transformed into protons on molecule **S** upon illumination of the sample with light. We would like to correlate the chemical shifts of the protons on molecule **I** with those on molecule **S**. Again for simplicity, the frequencies of the protons of **I** and of **S** are different and illumination of molecule **I** rapidly produces only molecule **S**. Assume that only **I** is present at the beginning of the experiment.

A simple approach is to “label” each individual proton of **I**, illuminate the sample, and then record the 1D spectrum of the resulting mixture. The “label” attached to the **I** proton is transferred to the corresponding proton in molecule **S**. A simple way to “label” an individual nucleus is to apply RF power at the frequency of the resonance in order to saturate the transition. A saturated resonance does not give rise to a peak in the spectrum and if an **I** proton is saturated followed by the transformation into a **S** proton by application of light, then the corresponding proton in **S** will be saturated and will not be observed in the spectrum (Fig. 4.1.1).

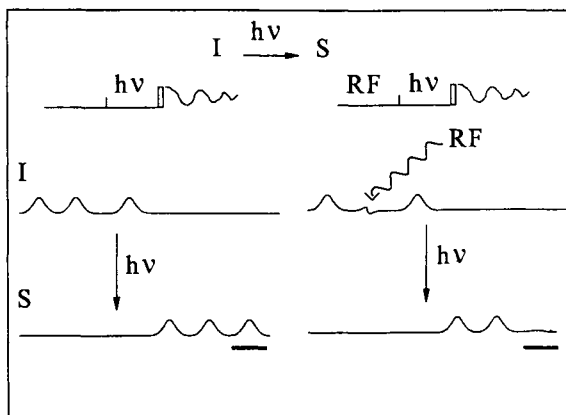


Fig. 4.1.1. One-dimensional saturation transfer experiment demonstrating the transfer of "labels" from one molecule, **I**, to another, **S**.

By systematically saturating each proton in a fresh sample of molecule **I**, followed by illumination, and then observation of the spectrum of the resulting mixture, each of the protons in molecule **I** could be correlated with the corresponding proton in molecule **S**. In general, this experiment consists of a preparation period (selective frequency saturation), a transfer or mixing period (illumination), followed by a detection period. A plot of the selective frequencies in **I** versus the corresponding transferred frequencies in **S** constitutes a two-dimensional map of the connectivities for the protons in **I** to the protons in **S**. If the detected spectrum is recorded as a pulsed FT spectrum then this experiment is collected as frequency (of selective saturation) versus time,  $S(f_1, t_2)$ , with  $f_1$  being the frequency of saturation and  $t_2$  being the detected free induction decay (FID). A Fourier transform on each individual experiment collected with a different selective frequency then generates the final frequency versus frequency,  $S(f_1, f_2)$  spectrum (Fig. 4.1.2).

This 2D spectrum could be collected in a mindless fashion by systematically moving the saturation frequency by a small, fixed amount across the width of the spectrum. At frequencies not resonant with any **I** protons, no effect would be seen on the spectrum, but at positions where the saturating frequency is resonant with an **I** spin then both the **I** spin and corresponding **S** spin will be saturated and be absent from the spectrum. Ideally this experiment would suffice for obtaining the correlations between **I** and **S** protons; in the real world, the saturating



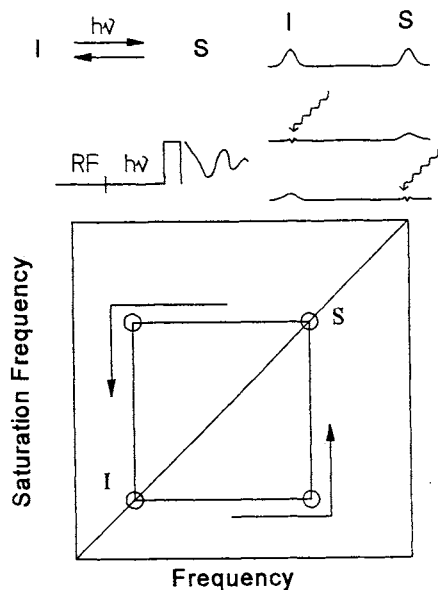


Fig. 4.1.2. Two-dimensional representation of a saturation transfer experiment.

frequency source has a finite bandwidth and a range of frequencies are excited when it is applied. For peaks that are close together in frequency, the saturating field will excite both spins and make the correlation to the S spins ambiguous.

## 4.2. Time-time two-dimensional spectroscopy

The Fourier transform defines a relationship between time and frequency. This relationship suggests that an experiment that is collected as a function of frequency can be substituted with an equivalent experiment collected as a function of time. An experiment equivalent to the saturation transfer experiment described can be collected as a time versus time,  $S(t_1, t_2)$ , two-dimensional spectrum. The details of the transfer are not exactly the same for the frequency-time and the time-time experiments but the results are essentially the same. In the frequency-time experiment, saturated magnetization is transferred, whereas in a time-time experiment the selective frequency information is passed as a time dependent difference magnetization. The pulse sequence for the two-dimensional EXSY experiment is shown in Eq. (4.1.1).



The preparation period for this experiment consists of allowing the spin system to come to thermal equilibrium in the static magnetic field followed by a  $90^\circ$ , which generates transverse magnetization. The preparation period is followed by the evolution delay  $t_1$ , then two  $90^\circ$  pulses separated by the mixing time,  $\tau_m$ , and a final acquisition time,  $t_2$ . The CFN for this EXSY experiment is shown in Fig. 4.1.3. The symbol **N** (Table 1.1), representing incoherent transfer of magnetization, is used in this CFN since the processes of chemical exchange and the nuclear Overhauser effect are formally equivalent.

The first  $90^\circ$  pulse tips the protons of **I** onto the  $-Y$  axis of the transverse plane, starting the evolution period (Fig. 4.1.4). In terms of product operators, this is described by

$$I_z = \pi/2 \hat{I}_x \Rightarrow -I_y. \quad (4.1.2)$$

Here **I** represents a proton spin in molecule **I**. If the receiver were turned on at this time the spectrum of **I** would be recorded; however, instead of detecting the signal during  $t_1$ , a series of spectra is recorded with  $t_1$  values that differ by a dwell time increment. The dwell time (DW) is related to the frequency range or sweep width (SW) that is to be covered by  $DW = 1/(2 \cdot SW)$ . During each of these  $t_1$  periods, the protons in **I** precess at their characteristic frequency,  $\omega_I$ . During a given  $t_1$  period, a particular proton will rotate in the transverse plane through an angle,  $\omega_I t_1$  (Eq. 4.1.3).

$$-I_y = \omega_I t_1 \hat{I}_z \Rightarrow -I_y \cos(\omega_I t_1) + I_x \sin(\omega_I t_1) \quad (4.1.3)$$

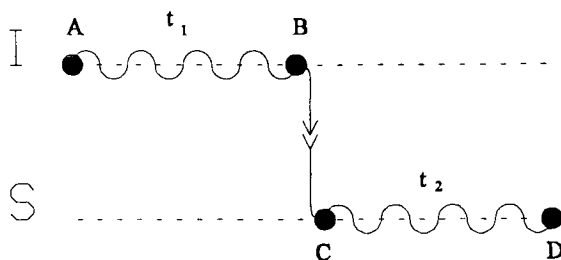


Fig. 4.1.3. CFN for two-dimensional exchange spectroscopy.

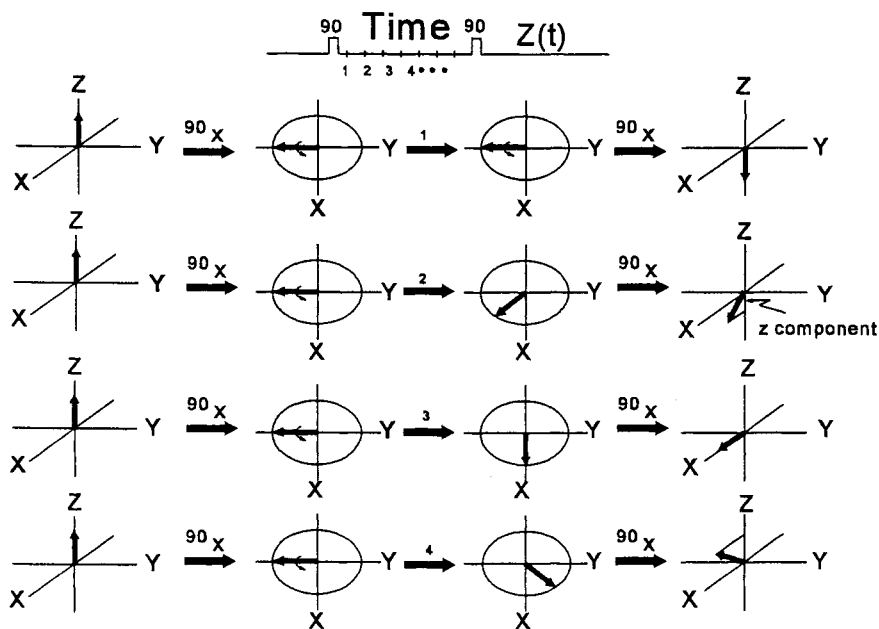


Fig. 4.1.4. Schematic representation of frequency labeling of Z magnetization by chemical shift.

The second  $90^\circ$  pulse along the X axis begins the mixing period by turning all of the Y components of the spin vectors into the Z axis (Eq. 4.1.4).

$$-I_y \cos(\omega_1 t_1) + I_x \sin(\omega_1 t_1) = \pi/2 \hat{I}_x \Rightarrow -I_z \cos(\omega_1 t_1) + I_x \sin(\omega_1 t_1) \quad (4.1.4)$$

The amount of Y component to be turned into the Z axis is directly related to the precession frequency (chemical shift). A proton that is on resonance with the carrier frequency does not precess in the rotating frame ( $\omega = 0$ ). Irrespective of the value of  $t_1$ , the vector will remain along the  $-Y$  axis. The second  $90^\circ$  pulse along the X axis will turn the vector into the  $-Z$  axis. For a proton that is not on resonance, the vector precesses at a frequency,  $\omega_1 t_1$ , in the rotating frame. At a value of  $t_1 = 0$ , the vector will lie along the  $-Y$  axis and will be turned along the  $-Z$  axis by the second  $90^\circ$  x pulse. As the value of  $t_1$  is incremented by the dwell time, the vector will precess in the XY plane by an angle  $\omega_1 t_1$ . The component of the vector along the  $-Y$  axis is given by  $\cos(\omega_1 t_1)$ . At  $t_1 =$

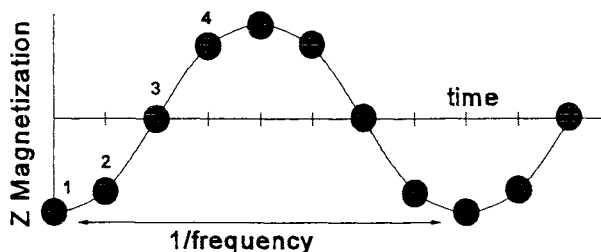


Fig. 4.1.5. Plot of the chemical shift modulated Z magnetization of Fig. 4.1.4.

$1/(4\omega_I/2\pi)$ , a vector precesses to lie along the X axis and the second  $90^\circ_x$  pulse does not effect it. A plot of the Z magnetization after the second  $90^\circ_x$  pulse versus  $t_1$  be the function  $\cos(\omega_I t_1)$  (Fig. 4.1.5). A Fourier transform of this time dependent function gives the frequencies,  $\omega_I$ , of the protons in molecule **I**.

During the delay  $\tau_m$ , the light is turned on to allow the reaction to proceed. An amount  $\lambda$  of **I** is converted to **S** (Eq. 4.1.5).

$$-I_z \cos(\omega_I t_1) + I_x \sin(\omega_I t_1) = hv \Rightarrow (1-\lambda) * [-I_z \cos(\omega_I t_1) + I_x \sin(\omega_I t_1)] + \lambda [-S_z \cos(\omega_I t_1) + S_x \sin(\omega_I t_1)] \quad (4.1.5)$$

The values of the Z magnetization for the protons of **I** that were established during the  $90^\circ-t_1-90^\circ$  period do not change during the application of light and thus the Z information is transferred from protons of **I** molecules to those of **S** molecules. The protons on **S** now have components along the Z axis that are a function of the frequency of the protons on **I**. Also during  $\tau_m$ , the X components of **I** protons that were not turned into the Z axis by the second  $90^\circ$  pulse precess and are converted into **S** proton transverse magnetization. These components are not desired and can be eliminated by phase cycling. We will defer the details of phase cycling to a later section. The elimination of the transverse components, in effect, leaves only Z magnetization at the end of  $\tau_m$  (Eq. 4.1.6).

$$(1-\lambda) * [-I_z \cos(\omega_I t_1)] + \lambda [-S_z \cos(\omega_I t_1)] \quad (4.1.6)$$

The final  $90^\circ$  pulse applied to both **I** and **S** spins, turns the Z magnetization along the Y axis and the FID is collected during the detection period (Eq. 4.1.7).

$$(1-\lambda)*[-\mathbf{I}_z \cos(\omega_I t_1)] + \lambda[-\mathbf{S}_z \cos(\omega_I t_1)] = \pi/2(\hat{\mathbf{I}}_x + \hat{\mathbf{S}}_x) \Rightarrow (1-\lambda)*[\mathbf{I}_y \cos(\omega_I t_1)] + \lambda[\mathbf{S}_y \cos(\omega_I t_1)] \quad (4.1.7)$$

The Y magnetization in the FID arises from two sources: 1). the protons on **I** molecules that did not react during  $\tau_m$  and 2). The protons on **S** that were created during  $\tau_m$ . Both types of protons retain the information about **I** proton frequencies that was stored as an amplitude modulation of the Z magnetization,  $\cos(\omega_I t_1)$ . During  $t_2$ , **I** and **S** protons precess at their characteristic frequencies,  $\omega_I$  and  $\omega_S$  (Eq. 4.1.8).

$$(1-\lambda) [\mathbf{I}_y \cos(\omega_I t_1)] + \lambda[\mathbf{S}_y \cos(\omega_I t_1)] = \omega_I t_2 \hat{\mathbf{I}}_x + \omega_S t_2 \hat{\mathbf{S}}_x \Rightarrow (1-\lambda) \cos(\omega_I t_1) * [\mathbf{I}_y \cos(\omega_I t_2) - \mathbf{I}_x \sin(\omega_I t_2)] + \lambda \cos(\omega_I t_1) * [\mathbf{S}_y \cos(\omega_S t_2) - \mathbf{S}_x \sin(\omega_S t_2)] \quad (4.1.8)$$

By collecting a series of experiments with an incrementation of  $t_1$  by the dwell time, a two-dimensional time–time spectrum,  $S(t_1, t_2)$ , is built up. A two-dimensional Fourier transform extracts the frequencies from the time dependent signal to generate a final spectrum  $S(f_1, f_2)$ .

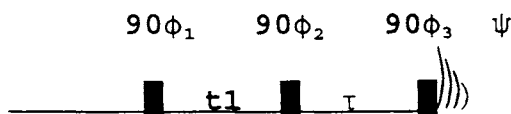
In 2D spectra, all peaks have two frequency coordinates corresponding to the frequency  $f_1$  that the spin experiences during the first time period  $t_1$  and the frequency  $f_2$ , which it has during the detection period  $t_2$ . In our example, an **I** spin in  $t_1$  that is transformed into a **S** spin during  $t_2$  will have the frequency coordinates,  $(\omega_I, \omega_S)$ . This type of peak is known as a *cross peak*. An **I** spin that remains an **I** spin during the experiment will have the same frequency coordinates in both dimensions  $(\omega_I, \omega_I)$ . This type of peak is known as a *diagonal peak*. In the simple experiment described here, the peaks that are observed are the diagonal peaks of **I** representing material that does not react during illumination and a cross peak at  $(\omega_I, \omega_S)$  that represents molecules of **I** that were transformed into **S** upon illumination. In a situation where the exchange proceeds in both directions simultaneously,  $\mathbf{I} \rightarrow \mathbf{S}$  and  $\mathbf{S} \rightarrow \mathbf{I}$ , a cross peak would also be observed at the symmetric coordinates  $(\omega_S, \omega_I)$  and a diagonal peak would appear at  $(\omega_S, \omega_S)$ .

Another type of peak observed in 2D nmr spectra arises from magnetization that relaxes during the evolution period. *Axial peaks* are found at zero frequency along  $f_1$ ,  $(0, \omega_I)$  and  $(0, \omega_S)$ . The relaxation of the magnetization during evolution produces Z magnetization without any frequency labeling. These components have not precessed during the  $t_1$

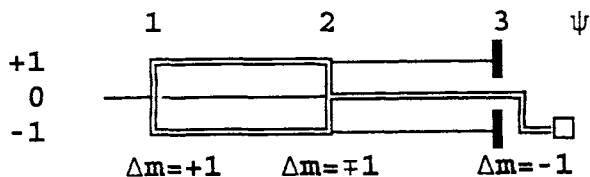
period and therefore appear to have zero frequency. These peaks can be removed by phase cycling.

### 4.3 Exchange spectroscopy (EXSY) phase cycling

The EXSY experiment consists of three  $90^\circ$  pulses; the first two pulses are separated by the  $t_1$  frequency labeling period and the second and third pulses are separated by the mixing time during which magnetization is exchanged among the spins.



From knowledge of the coherence order at each step of the pulse sequence a coherence transfer pathway can be constructed.



The first  $90^\circ$  pulse excites equilibrium magnetization to coherence order  $\pm 1$ . From thermal equilibrium this is the only possibility with a single pulse. Of course some magnetization may remain along the Z axis due to pulse misset or imperfections. The change in coherence order is the parameter that is important in designing a phase cycling routine, for pulse 1:

$$\Delta m = \pm 1.$$

Following the procedure of Section 3.6, for pulse 1 we list and select the desired coherence order changes.

$$\phi_1: (-1 \ 0) + 1.$$

The  $-1$  value was chosen as the desired value ( $+1$  could have been chosen with the same results). The number of members is:

$$N=2.$$

Computing the phases for pulse 1 from Eq. (3.6.1):

$$\phi_1 = 0, \pi$$

The role of the second pulse is to generate Z magnetization from the precessing transverse magnetization. Any remaining transverse magnetization after this pulse is undesirable. The phase cycle is designed to only pass zero order coherence, such as Z, during the mixing period. The coherence transfer pathway is drawn such that the  $\pm 1$  coherences are transferred to the 0 level. Note that since the magnetization is not at thermal equilibrium, the  $\pm 1$  selection rules are dropped. This is not important in the isolated spin system described here, but in a spin system with  $\mathbf{n}$  coupled spins the coherence levels  $+\mathbf{n}, -\mathbf{n}$  can become populated. These higher order coherences must be considered in the overall phase cycle.

Following the same procedure as for pulse 1, the list of  $\Delta m$  values with the proper marking is

$$\phi_2: -1(0 \ +1).$$

Note that to go from order  $-1$  to order  $0$ ,  $\Delta m = +1$ . Again  $N=2$  and the phases are calculated as before

$$\phi_2 = 0, \pi$$

For the third pulse, the coherence is transferred to order  $-1$  that is then detected by the receiver. Quadrature images arise from magnetization that transfers to the  $+1$  level. As above, from the list of  $\Delta m$  values select the subset of coherences:

$$\phi_3: (+2 \ +1 \ 0 \ \underline{-1}) -2$$

Since there is not another value above  $+2$  that we want, we move to the farthest point on the list. We see that  $N=4$ . Unlike in the CYCLOPS phase cycle, Section 3.9, where we needed to add a  $+2$  coherence order level to obtain  $90^\circ$  phase shifts, the  $\pm 2$  coherence order changes are not fictitious in this step in the phase cycle.

The phases are:

$$\phi_3 = 0, \pi/2, \pi, 3\pi/2$$

Using the formula (Eq. 3.6.2), the receiver phase can be calculated. Each pulse is systematically and independently cycled through the proper phases. The receiver is phase cycled to cause constructive interference for signals that have taken the “right” pathway and to cause

destructive interference for those that have not.

Let us calculate the phase for one step in the phase cycle. Let  $\phi_1 = \pi$ ;  $\phi_2 = \pi$ ; and  $\phi_3 = 3\pi/2$ . The receiver phase,  $\psi$ , is given by

$$\psi = -[(-1)(\pi) + (+1)(\pi) + (-1)(3\pi/2)] = 3\pi/2.$$

Notice that the path that was followed was  $\Delta m_1 = -1$ ;  $\Delta m_2 = +1$ ;  $\Delta m_3 = -1$ . The same phase cycle is obtained by following the path:  $\Delta m_1 = +1$ ;  $\Delta m_2 = -1$ ;  $\Delta m_3 = -1$ . Finally, the phase table for the EXSY experiment follows.

$$\phi_1: 8(0, \pi)$$

$$\phi_2: 4(0, 0, \pi, \pi)$$

$$\phi_3: 4(0), 4(\pi/2), 4(\pi), 4(3\pi/2)$$

$$\psi: (0, \pi, \pi, 0) (\pi/2, 3\pi/2, 3\pi/2, \pi/2) (\pi, 0, 0, \pi) (3\pi/2, \pi/2, \pi/2, 3\pi/2)$$

In this notation, 4(0) represents four steps of 0 phase, 8(0,  $\pi$ ) represents eight pairs of 0 and  $\pi$  phases, etc. This experiment is carried out by collecting 16 individual acquisitions and combining them in the computer memory.

Both + and - pathways must be taken during the  $t_1$  evolution period to obtain spectra with pure phase. By allowing both pathways, the frequency is transferred by amplitude modulation that provides for pure phase spectra. If only one pathway is selected the frequency is transferred by phase modulation and the resulting line shapes are an inseparable mixture of absorptive and dispersive peaks. This is not a good situation for obtaining high resolution spectra. As well, the choice of a single pathway, as in "magnitude COSY," throws away ~30% of the signal amplitude, leading to poor signal-to-noise.

## 5. COHERENCE TRANSFER

In introducing coherence transfer, it is easy to get bogged down in a large number of product operator terms and obscure terminology. Here I present a simplistic approach to coherence transfer. I apologize to those who do not like semi-classical, hand-waving descriptions of quantum mechanics. The experiment I will use here to introduce the subject is an old experiment that has little direct practicality in modern multi-dimensional NMR. However, it serves as an important "building block"



in many multidimensional heteronuclear experiments. The experiment correlates the chemical shifts of a proton and its attached carbon by transferring frequency labeled proton magnetization to a directly bonded carbon and detecting the carbon signal. Variants of this experiment are very important in many current multidimensional, heteronuclear experiments.

### 5.1. INEPT type transfer [22]

The one-bond coupling constant between a  $^1\text{H}$  and  $^{13}\text{C}$ ,  $^1J_{\text{HC}}$ , is about 140 Hz for aliphatic ( $\text{sp}^3$ ) type carbons. The semiclassical picture of an ensemble of  $^1\text{H}$  nuclei attached to  $^{13}\text{C}$  nuclei is a mixture of molecules with about half of the protons attached to a  $^{13}\text{C}$  nucleus with spin "down,"  $\beta$ , and the remainder attached to nuclei with spin "up",  $\alpha$ . The attached  $\alpha$  and  $\beta$   $^{13}\text{C}$  nuclei give rise to different local magnetic fields at the  $^1\text{H}$  nucleus. One  $^{13}\text{C}$  spin state adds to the external magnetic field and increases the frequency of resonance, while the other state decreases the magnetic field and the corresponding resonance frequency. The one-dimensional  $^1\text{H}$  NMR spectrum of a molecule containing a  $^1\text{H}$  attached to a  $^{13}\text{C}$  (such as  $^{13}\text{C}^1\text{HCl}_3$  ignoring the quadrupolar spins on the Cl nucleus) consists of a doublet centered at the  $^1\text{H}$  frequency and split by the frequency of the coupling constant  $J_{\text{HC}}$ . The two peaks arise from the  $\alpha$ -labeled  $^1\text{H}$  and the  $\beta$ -labeled  $^1\text{H}$ ,

$$\langle\alpha\rangle[-\mathbf{H}_y] + \langle\beta\rangle[-\mathbf{H}_y] \quad (5.1.1)$$

where the  $\langle\alpha\rangle$  and  $\langle\beta\rangle$  can be viewed as weighting coefficients having values determined from the  $^{13}\text{C}$  Boltzmann distribution.

Upon applying a  $90^\circ$  pulse to the mixture of longitudinal  $^1\text{H}$  spins with polarization  $\epsilon$  at thermal equilibrium (Fig. 5.1.1), we have,

$$\epsilon(\langle\alpha\rangle\mathbf{H}_z + \langle\beta\rangle\mathbf{H}_z) = \pi/2_x \hat{\mathbf{H}}_x \Rightarrow \epsilon(\langle\alpha\rangle[-\mathbf{H}_y] + \langle\beta\rangle[-\mathbf{H}_y]) \quad (5.1.2)$$

The amplitude of the induced transverse magnetization is  $\epsilon$ . If the proton resonance, in the absence of coupling, is on-resonance (i.e. zero frequency in the rotating frame), then the two  $\mathbf{H}_y$  vectors will precess in opposite directions, each precessing away from the  $-Y$  axis due to the local magnetic fields generated by the opposite spin states of the attached  $^{13}\text{C}$  nucleus. The frequency difference between the two vectors is equal to the coupling constant. At a time equal to  $[1/(2*J_{\text{CH}})]$ , the two

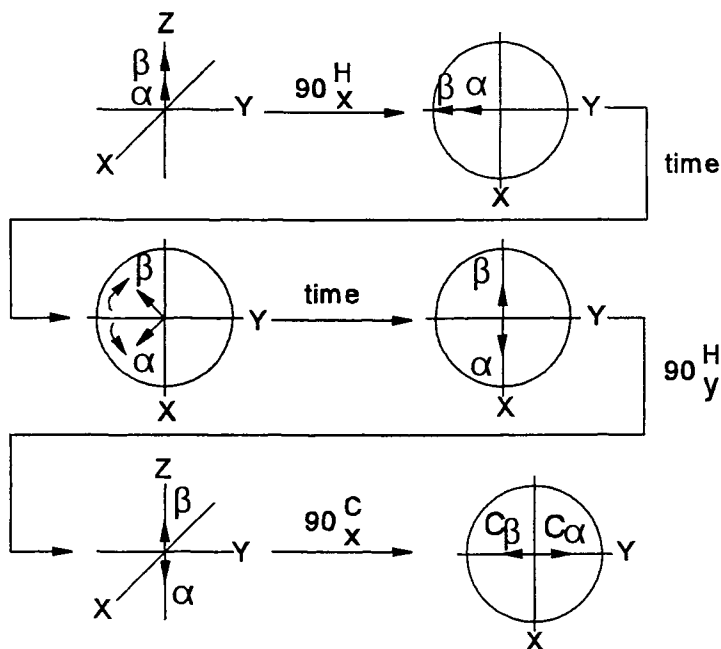


Fig. 5.1.1. Vector representation of INEPT transfer for a coupled  $^1\text{H}$ - $^{13}\text{C}$  spin system.

spin vectors will have precessed into opposite positions along the X coordinate axis.

$$\varepsilon\{\langle\alpha\rangle[-\mathbf{H}_y] + \langle\beta\rangle[-\mathbf{H}_y]\} = (1/(2J_{\text{CH}})) \Rightarrow \varepsilon\{\langle\alpha\rangle[\mathbf{H}_x] + \langle\beta\rangle[-\mathbf{H}_x]\} \quad (5.1.3)$$

Equation (5.1.3) can be rearranged to

$$\varepsilon\{\langle\alpha\rangle\mathbf{H}_x - \langle\beta\rangle\mathbf{H}_x\} = \varepsilon\{\mathbf{H}_x(\langle\alpha\rangle - \langle\beta\rangle)\}, \quad (5.1.4)$$

but the quantity  $(\langle\alpha\rangle - \langle\beta\rangle)$  is proportional to the longitudinal magnetization of the coupled spin  $\mathbf{C}_z$  [23]. By substituting  $\mathbf{C}_z$  for  $(\langle\alpha\rangle - \langle\beta\rangle)$  one obtains

$$\varepsilon\{\mathbf{H}_y\} = (1/(2J_{\text{CH}})) \Rightarrow \{2\mathbf{H}_x\mathbf{C}_z\} \quad (5.1.5)$$

where the "2" preceding the operator is a normalization factor. The state  $2\mathbf{H}_x\mathbf{C}_z$  is an antiphase state;  $\mathbf{H}_x$  magnetization is antiphase with re-

spect to  $\mathbf{C}$ . As evolution of the spins continues, the two proton vectors align along the  $\mathbf{H}_y$  axis at time  $1/J_{\text{CH}}$ , then after a total period of  $3/(2J_{\text{CH}})$  the antiphase state  $-2\mathbf{H}_x\mathbf{C}_z$  will form, and finally  $-\mathbf{H}_y$  will return at a total time of  $2/(J_{\text{CH}})$ . The time dependence is represented as

$$-\mathbf{H}_y = \pi J_{\text{CH}} t \hat{2}\mathbf{H}_z\hat{\mathbf{C}}_z \Rightarrow -\mathbf{H}_y \cos(\pi J_{\text{CH}} t) + 2\mathbf{H}_x\mathbf{C}_z \sin(\pi J_{\text{CH}} t). \quad (5.1.6)$$

Transverse magnetization oscillates between inphase,  $-\mathbf{H}_y$ , and antiphase magnetization,  $2\mathbf{H}_x\mathbf{C}_z$ . Note that the rotation operator is  $=\pi J_{\text{CH}} t \hat{2}\mathbf{H}_z\hat{\mathbf{C}}_z \Rightarrow$  for the coupling interaction.

If the spin system is in the  $2\mathbf{H}_x\mathbf{C}_z$  state and a  $90^\circ$  pulse is applied to the  $^1\text{H}$  nuclei, rotates the antiphase vectors into the  $\pm Z$  axis (Fig. 5.1.1). This is described in product operators as the single state of longitudinal two-spin order,  $2\mathbf{H}_z\mathbf{C}_z$ .

$$\varepsilon\{2\mathbf{H}_x\mathbf{C}_z\} = \pi/2 \hat{2}\mathbf{H}_y \Rightarrow \varepsilon\{2\mathbf{H}_z\mathbf{C}_z\} \quad (5.1.7)$$

Semi-classically, one vector can be associated with a Boltzmann equilibrium population while the other has an inverted population. To better visualize this process, Figure 5.1.2 shows the energy level diagrams for a two-spin  $^1\text{H}$ - $^{13}\text{C}$  spin system in various polarization states. Schematic one-dimensional spectra in each panel of Fig. 5.1.2 represent the signals that would be obtained by applying a  $90^\circ$  pulse to the various polarization states. Figure 5.1.2A is the energy level diagram for the system that has just had the magnetic field applied. The energy levels have been established, but there has not been time to establish thermal equilibrium. In Figure 5.1.2B, the system has reached Boltzmann equilibrium through the processes of spin-lattice relaxation. The polarization across the  $^1\text{H}$  transitions is four times that of the  $^{13}\text{C}$  spins, reflecting the 4:1 ratio of the  $^1\text{H}$  magnetogyric ratio to that of  $^{13}\text{C}$ . Panel C of Fig. 5.1.2 represents a state in which the  $^1\text{H}$  spin polarization across one transition has been inverted. A spectrum of the  $^1\text{H}$  spins would show one peak of the doublet inverted with respect to the other. Since the  $^{13}\text{C}$  spins share the same energy levels, the population differences across the  $^{13}\text{C}$  transitions are also perturbed. In Fig. 5.1.2D, where the  $^{13}\text{C}$  spins are saturated, the polarization across the  $^{13}\text{C}$  transitions is identical to those across the  $^1\text{H}$  transitions. If a  $90^\circ$  pulse is applied to the  $^{13}\text{C}$  spins in the states described in Fig. 5.1.2C or Fig. 5.1.2D, the transverse  $^{13}\text{C}$  coherence, which is antiphase to  $^1\text{H}$ , reflects the perturbed populations. A comparison of Fig. 5.1.2B with Fig. 5.1.2C

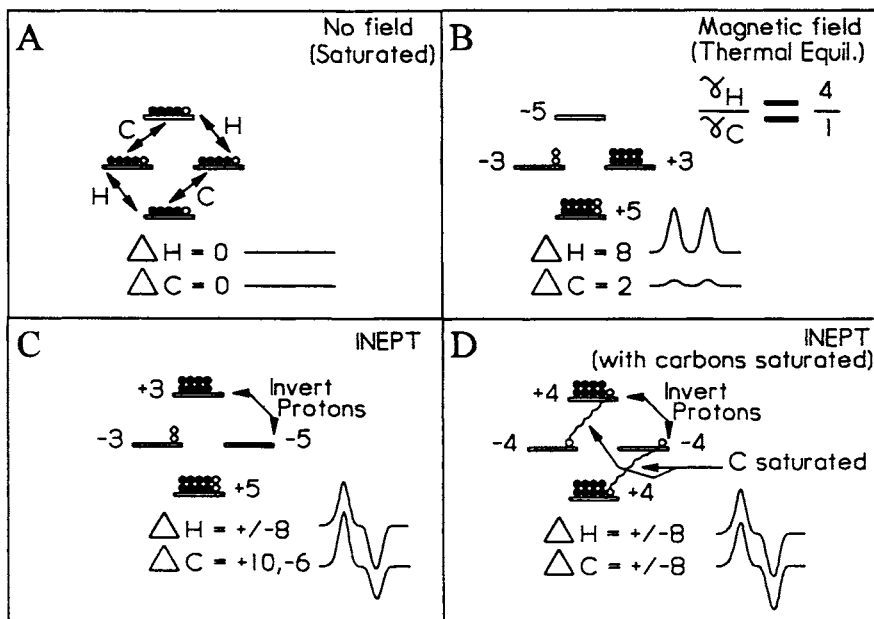


Fig. 5.1.2. Energy level diagrams of a two-spin  $^1\text{H}$ - $^{13}\text{C}$  system showing various polarizations.

and Fig. 5.1.2D, shows the increased signal amplitude of the  $^{13}\text{C}$  spins due to polarization transfer from the  $^1\text{H}$  spins.

$$\varepsilon\{2\text{H}_z\text{C}_z\} = \pi/2 \hat{\text{C}}_x \Rightarrow \varepsilon\{-2\text{H}_z\text{C}_y\} \quad (5.1.8)$$

Like the antiphase  $^1\text{H}$  coherence, this coherence evolves due to scalar coupling into an inphase  $^{13}\text{C}$  resonance in a time of  $1/(2J_{\text{CH}})$ .

$$\varepsilon\{-2\text{H}_z\text{C}_y\} = \pi J_{\text{CH}} (1/(2J_{\text{CH}})) \hat{\text{C}}_x \hat{\text{C}}_z \Rightarrow \varepsilon\{\text{C}_x\} \quad (5.1.9)$$

The initial polarization of the  $^1\text{H}$  spins,  $\varepsilon$ , has been transferred to the  $^{13}\text{C}$  spins by this sequence (Eq. 5.1.9).

As in the EXSY experiment, the  $^1\text{H}$  spins can be frequency labeled by point-by-point sampling of a  $t_1$  evolution period during which the  $^1\text{H}$  spins precess under the chemical shift Hamiltonian. The chemical shift is represented as a modulation of the transferred polarization and in turn a signal amplitude of the  $^{13}\text{C}$  spins. The amplitude modulation of

the  $^{13}\text{C}$  spins encodes the  $^1\text{H}$  chemical shift frequency obtained during the  $t_1$  period.

One can now design a 2D experiment using this mechanism of coherence transfer. Without detailing the exact pulse sequence, this experiment can be described with a coherence flow network (CFN). The CFN for this experiment is shown in Fig. 5.1.3. The letters in brackets, e.g.  $\langle A, B \rangle$ , correspond to the various sections of the CFN. The spins are assumed to be in thermal equilibrium at state  $\langle A \rangle$ . A  $90^\circ$  pulse transfers the  $^1\text{H}$  magnetization into the transverse plane. The section  $\langle A, B \rangle$  represents the chemical shift precession of the  $^1\text{H}$  spins, this is the  $t_1$  period that will be sampled point-by-point in time. The absence of any modifier during this period shows that the chemical shift precession is not modulated by any coupling, i.e. the carbons are decoupled from the protons. At state B, the  $^1\text{H}$  spins have been frequency labeled. During the  $\langle B, C \rangle$  period, active coupling (Table 1.1A) between  $^1\text{H}$  and  $^{13}\text{C}$  evolves without any chemical shift precession (Table 1.1T). At state  $\langle C \rangle$ , the proton spin is antiphase with respect to the carbon spin. Coherence transfer from transverse  $^1\text{H}$  to transverse  $^{13}\text{C}$  occurs along  $\langle C, D \rangle$ . Active coupling between the spins evolves during  $\langle D, E \rangle$  refocusing the antiphase state at  $\langle D \rangle$  to an inphase  $^{13}\text{C}$  spin at  $\langle E \rangle$ . The total period  $\langle B, E \rangle$  represents a refocused INEPT transfer. Finally, the chemical shift frequency of the  $^{13}\text{C}$  spin without coupling to  $^1\text{H}$  is detected during  $\langle E, F \rangle$ . Two-dimensional Fourier transform of the resulting  $S(t_1, t_2)$  data yields a spectrum with a peak arising at the chemical shift coordinates of the  $^1\text{H}$  and its coupled  $^{13}\text{C}$ . Figure 5.1.4 is a schematic 2D  $^{13}\text{C}\{^1\text{H}\}$  spectrum that might be collected using an experiment of this type. Along the top of Fig. 5.1.4 is the (schematic) 1D  $^{13}\text{C}$  spectrum and along the side is the corresponding 1D  $^1\text{H}$  spectrum. At the intersection of pairs of  $^1\text{H}$ – $^{13}\text{C}$  frequencies are cross peaks establishing that the two nuclei are directly coupled.

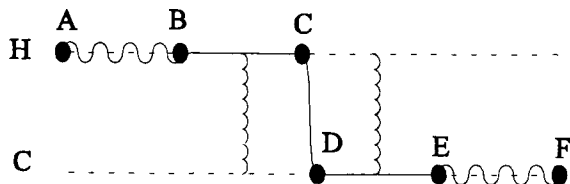


Fig. 5.1.3. CFN for a  $^{13}\text{C}\{^1\text{H}\}$  chemical shift correlation experiment.

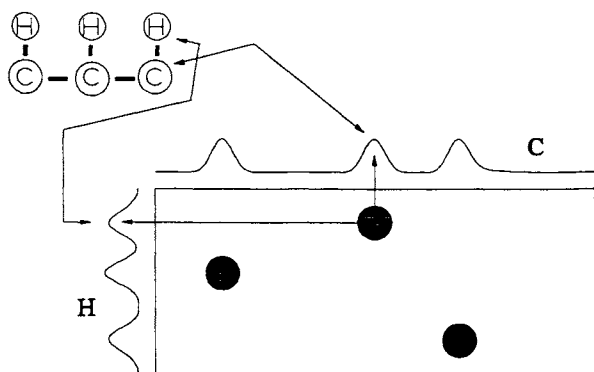


Fig. 5.1.4. Schematic  $^{13}\text{C}\{^1\text{H}\}$  heteronuclear chemical shift correlation spectrum.

## 6. HOMONUCLEAR CORRELATED SPECTROSCOPY (COSY)

Homonuclear Correlated spectroscopy (COSY) was the first 2D experiment introduced. Jeener [2] proposed the experiment in 1971 and Aue, Barthodi, and Ernst [3] analyzed and experimentally demonstrated the technique in their seminal paper in 1976. Several modifications to COSY quickly followed and the experiment was applied to many systems.

### 6.1. COSY

The COSY pulse sequence consists of two pulses separated by an incremented  $t_1$  delay. The spins are allowed to come to thermal equilibrium (or a steady state) before applying the first  $90^\circ$  pulse, which generates transverse magnetization at  $\langle A \rangle$  in the COSY CFN (Fig. 6.1.1). During  $\langle A, B \rangle$ , spin  $I$  precesses at the  $I$  spin frequency and simultaneously evolves via coupling into antiphase  $2I_z S_z$  at  $\langle B \rangle$ . Coherence transfer from  $I$  to  $S$ ,  $\langle B, C \rangle$ , is effected by the second  $90^\circ$  pulse on both spins. The  $S$  spin then evolves under both  $S$  chemical shift and coupling to  $I$  during detection,  $\langle C, D \rangle$ .

Cross peaks in COSY appear at the intersection of the frequencies of the scalar coupled spins. In this example, a cross peak would occur at the  $I$  spin frequency along  $\omega_1 I$  and at the  $S$  spin frequency along  $\omega_2 S$ ,  $(\omega_1 I, \omega_2 S)$ . Since the  $S$  spin also serves as a source of magnetization, there will also be a cross peak representing  $S$  coherence that is trans-

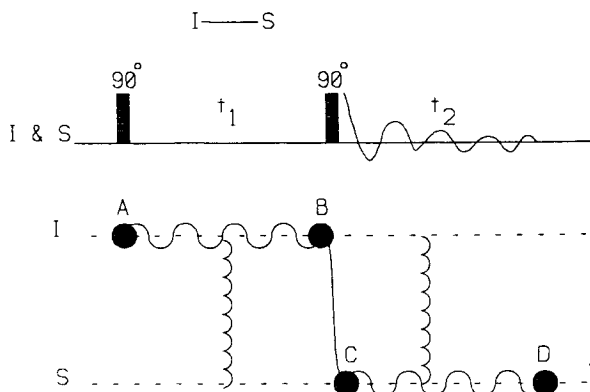


Fig. 6.1.1. COSY pulse sequence and CFN diagram.

ferred to **I** coherence at  $(\omega_1\mathbf{S}, \omega_2\mathbf{I})$ . The active coupling that occurs during  $\langle A, B \rangle$  and  $\langle C, D \rangle$  gives rise to multiplet peaks that are antiphase (see Fig. 6.2.1).

Diagonal peaks in a COSY spectrum arise from magnetization that remains on the same spin after the second  $90^\circ$  pulse. The frequencies are identical for both  $t_1$  and  $t_2$  periods and peaks are found at  $(\omega_1\mathbf{I}, \omega_2\mathbf{I})$  and  $(\omega_1\mathbf{S}, \omega_2\mathbf{S})$ . Figure 6.1.2 contains the CFN for the diagonal peaks of the COSY spectrum. Note that, unlike the case for the cross peaks, there is only passive coupling during both  $t_1$  and  $t_2$ . Passive coupling results in multiplet peaks being inphase in the final spectrum. The CFN does not explicitly indicate the relative phase of the cross and diagonal peaks; below the detailed analysis shows that the cross peaks and diagonal peaks are  $90^\circ$  out of phase (see Fig. 6.2.2).

*Preparation:*

$$\mathbf{I}_z = \pi/2 \hat{\mathbf{I}}_x \Rightarrow -\mathbf{I}_y \quad (6.1.1)$$

*Evolution ( $t_1$ ):*

$$= \omega_1 t_1 \hat{\mathbf{I}}_z \Rightarrow -\mathbf{I}_y \cos \omega_1 t_1 + \mathbf{I}_x \sin \omega_1 t_1 \quad (6.1.2)$$

$$= \pi J_{IS} t_1 2 \hat{\mathbf{I}}_z \hat{\mathbf{S}}_z \Rightarrow (-\mathbf{I}_y \cos \pi J_{IS} t_1 + 2 \mathbf{I}_x \mathbf{S}_z \sin \pi J_{IS} t_1) \cos \omega_1 t_1 \\ + (\mathbf{I}_x \cos \pi J_{IS} t_1 + 2 \mathbf{I}_y \mathbf{S}_z \sin \pi J_{IS} t_1) \sin \omega_1 t_1 \quad (6.1.3)$$

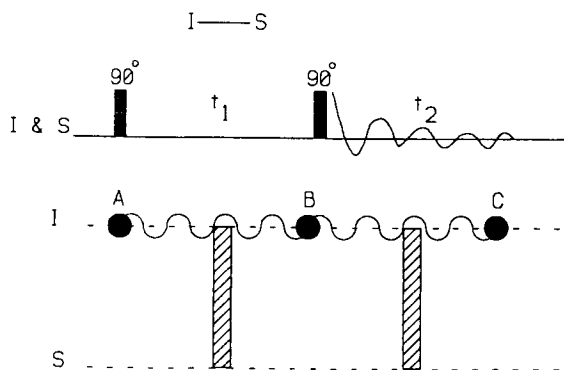


Fig. 6.1.2. CFN for COSY diagonal peaks.

*Mixing:*

$$\begin{aligned}
 =\pi/2(\hat{\mathbf{I}}_x + \hat{\mathbf{S}}_x) \Rightarrow & (-\mathbf{I}_x \cos \pi J_{IS} t_1 - 2\mathbf{I}_x \mathbf{S}_y \sin \pi J_{IS} t_1) \cos \omega_I t_1 \\
 & + (\mathbf{I}_x \cos \pi J_{IS} t_1 - 2\mathbf{I}_z \mathbf{S}_y \sin \pi J_{IS} t_1) \sin \omega_I t_1 \quad (6.1.4)
 \end{aligned}$$

During the detection period,  $t_2$ , the spin system evolves under chemical shift and coupling operators. The receiver is turned on and the signals are detected.

If all of the terms are retained, the calculation becomes very cumbersome due to the large number of terms that are generated. A great simplification of the product operator calculations is achieved by only calculating the evolution of the observable signals during the detection period. Only  $\mathbf{I}_x$ ,  $\mathbf{I}_y$ ,  $\mathbf{S}_x$ , and  $\mathbf{S}_y$  give directly observable signals in the NMR spectrometer. The pertinent terms at the beginning of the detection period are the direct observables and any terms that will evolve into observables due to coupling (e.g.,  $2\mathbf{I}_z \mathbf{S}_y$ ). All other coherences are not observable. During the detection period of the COSY experiment there are 13 terms that evolve. All of the directly observable terms arise from the  $\mathbf{I}_x$  and  $2\mathbf{I}_z \mathbf{S}_y$  terms created during the mixing period. The  $\mathbf{I}_x$  term evolves by chemical shift into  $\mathbf{I}_y$ , and the  $2\mathbf{I}_z \mathbf{S}_y$  term evolves through coupling into  $\mathbf{S}_x$  that further evolves into  $\mathbf{S}_y$  by chemical shift. The interesting terms present at the beginning of the detection period are:

$$\mathbf{I}_x \cos \pi J_{IS} t_1 \sin \omega_I t_1 \quad (6.1.5)$$



and

$$-2\mathbf{I}_z\mathbf{S}_y \sin \pi J_{\text{IS}}t_1 \sin \omega_{\text{I}}t_1 \quad (6.1.6)$$

These signals will further evolve under  $J_{\text{IS}}$  coupling and chemical shift during  $t_2$  into

$$\mathbf{I}_x \exp(i\omega_{\text{I}}t_2) \cos \pi J_{\text{IS}}t_2 \cos \pi J_{\text{IS}}t_1 \sin \omega_{\text{I}}t_1 \quad (6.1.7)$$

and

$$\mathbf{S}_x \exp(i\omega_{\text{S}}t_2) \sin \pi J_{\text{IS}}t_2 \sin \pi J_{\text{IS}}t_1 \sin \omega_{\text{I}}t_1. \quad (6.1.8)$$

## 6.2. Interpretation of COSY spectra

The signals that have the frequency of the **I** spin during both  $t_1$  and  $t_2$  are diagonal peaks (Eq. 6.1.5). The **S** signals, which precess at the **S** frequency during  $t_2$ , are frequency labeled during  $t_1$  with the frequency of the **I** spin; these are the cross peaks (Eq. 6.1.6). There are two significant features of these different signals: 1). the coupling modulation of the cross peak signals in both  $t_1$  and  $t_2$  is a sine function, whereas for the diagonal peaks the modulation is a cosine function; and 2). there is a  $90^\circ$  phase difference between the  $-\mathbf{I}_y$  and  $2\mathbf{I}_z\mathbf{S}_x$  terms at the beginning of the acquisition (Eqs. 6.1.5 and 6.1.6). The first feature means that the cross peak multiplets are antiphase in both dimensions (Fig. 6.2.1) whereas the diagonal multiplets are inphase in both dimensions (Fig. 6.2.2.). The second feature means that when the cross-peaks are phased to absorption mode, as in Fig. 6.2.1, then the diagonal peaks have a dispersive line shape shown in Fig. 6.2.2. The dispersive nature of the diagonal peaks cause perturbations over a wide area around the peaks. Notice the extent of the baseline distortions around the diagonal peaks (Fig. 6.2.2) as compared to the flat baseline around the cross peak in (Fig. 6.2.1). For peaks with linewidths on the order of the coupling constant, the positive part of the antiphase line partially cancels the negative part reducing the intensity of the cross peak. The diagonal peaks, on the other hand, are inphase and the multiplets constructively interfere. This combined with the fact that the diagonal peaks are phased to a dispersive lineshape, makes the diagonal peaks large. Cross peaks of coupled spins having similar chemical shifts are close to the diagonal and are difficult to observe since they are swamped by the

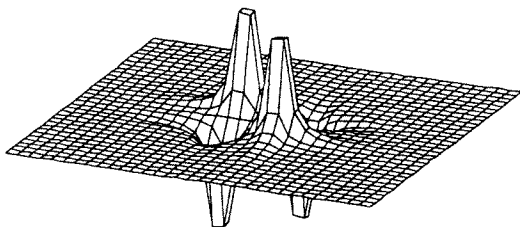


Fig. 6.2.1. Simulated antiphase, absorptive COSY-type cross peak. the linewidth used in the simulation was 10% of the coupling constant.

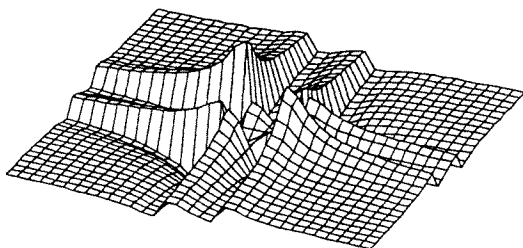


Fig. 6.2.2. Simulated diagonal peak from a COSY experiment. The linewidth to coupling constant ratio was the same as in Fig. 6.2.2.

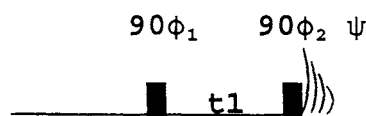
large diagonal peaks. The combination of these features creates undesirable spectral characteristics in the COSY spectrum. Application of a heavy filter function (e.g. unshifted sinebell) reduces the diagonal amplitude with respect to the cross peaks while distorting the line-shape. Mueller [24] has used a time domain convolution difference filter to reduce the diagonal peaks of a COSY spectrum with little disturbance of cross peaks.

A common misconception held by novices is that the cancellation of cross peaks due to linewidths that are of the same order or larger than the coupling constant can be overcome by using the “magnitude” mode instead of the “phase sensitive” mode of collecting COSY data. The confusion comes from the differences in the appearance of the spectra obtained by the two methods. Peaks in magnitude mode spectra have only positive peaks since during *processing* the square root of the sum of the squares (the magnitude) of the absorption and corresponding dispersion peaks is obtained. Since it appears that all of the cross peak components have the same algebraic sign, they *apparently* do not cancel

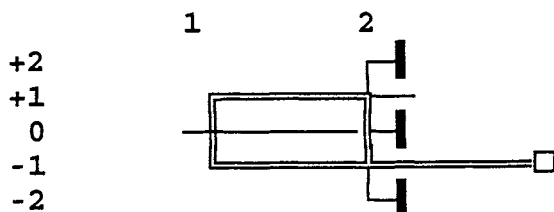
as they would in a phase sensitive spectrum where the spectrum is displayed as positive and negative absorption mode peaks. However, the sensitivity of the phase sensitive spectra is theoretically greater than that of the magnitude spectrum by a square root of two. The reason that the magnitude calculation is performed in magnitude mode spectra is that the inherent lineshape of the cross peaks is phase twisted, i.e. the peaks contain an inseparable mixture of both *antiphase* absorptive and *antiphase* dispersive line shapes. The resolution afforded by phase-twisted line shapes is not as good as that of the pure absorption line shape. The data processing of the magnitude applies a sinebell or other digital filter that induces approximate symmetry on the time domain data. Fourier transform of the data followed by a magnitude calculation gives approximately absorptive lineshapes while throwing away the phase information. Any cancellation in the cross peak has already occurred during the generation of the antiphase state when the data was collected. The extra square-root-of-two sensitivity in the phase sensitive spectrum comes from the additional coherence transfer pathway that is allowed by the phase cycling.

### 6.3. COSY phase cycling

Correlation spectroscopy (COSY):



COSY Coherence Transfer Pathway:



The first pulse excites longitudinal to transverse magnetization. The desired coherence order changes are then

$$\Delta m_1 = \pm 1$$

The second pulse is the mixing pulse; all possible coherence order changes can occur. The desired changes are those that cause transfer from  $+1 \rightarrow -1$  coherence and  $-1 \rightarrow -1$  coherence (Eq. 6.1.4), specifically,

$$2I_yS_z \Rightarrow 2I_zS_y$$

and

$$I_x \Rightarrow I_x$$

These changes represent transfers of  $\Delta m_2 = 0, -2$ .

Following the rules for phase cycling (Section 3.6), we obtain the phase cycles for  $\phi_1$ ,  $\phi_2$ , and  $\psi$ .

$\phi_1$ :	$\phi_2$ :
+1 (0 -1)	+3 <b>+2</b> +1 <b>0</b> (-1 <b>-2</b> ) -3
N=2	N=2
$\phi_1 = 2\pi k/2$ k=0,1	$\phi_2 = 2\pi k/2$ k=0,1
$\phi_1 = 0, \pi$	$\phi_2 = 0, \pi$

Note that for  $\phi_2$  both  $\Delta m=0$  and  $+2$  are in bold. These order changes are also allowed, as are all order changes for  $\Delta m = \Delta m_{\text{sel}} \pm n*N$  with n an integer.

For the receiver phase:

$$\psi = 0, \pi, 0, \pi$$

e.g., for step 4 in the table below, the receiver phase is calculated for the  $-1, 0$  pathway as:

$$\psi_4 = -[(-1)(\pi) + (0)(\pi)] = \pi$$

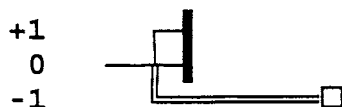
or for  $+1, -2$  pathway

$$\psi_4 = -[(+1)(\pi) + (-2)(\pi)] = \pi$$

This leads to a very simple phase table for COSY experiment.

$$\begin{aligned}\phi_1 &= 2(0 \pi) \\ \phi_2 &= 2(0) 2(\pi) \\ \psi &= 2(0 \pi)\end{aligned}$$

Since this phase cycle permits  $\Delta m = +2$  transfer at the second pulse, the resulting  $-1 \Rightarrow +1$  transfer can lead to quadrature images. To restrict the CTP further one can force the coherence to start at level 0 and end at level  $-1$ . In this CTP, the entire COSY pulse sequence is used as a black box that generates  $\Delta m = -1$  transfer from thermal equilibrium. The equivalent overall CTP is:



Analyzing this overall CTP we obtain

$\phi_c$ :

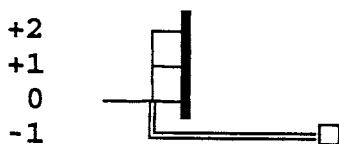
$$\Delta m_1 = -1$$

$$(-1 \ 0 \ 1)$$

$$N=3$$

$$\phi_c = 0, 2\pi/3, 4\pi/3$$

This CTP leads to  $120^\circ$  phase shifts, which are not available on all spectrometers. It is always permitted to eliminate coherence pathways that do not exist, so to use  $90^\circ$  phase shifts, available on most spectrometers, we add a fictitious transfer of  $+\Delta 2$  from the 0 level. Note that from thermal equilibrium this is a forbidden transition.



Again we calculate the phase cycle

$\phi_c$ :

$$\Delta m_1 = -1$$

$$(-1 \ 0 \ 1 \ 2)$$

$$N=4$$

$$\phi_c = 0, \pi/2, \pi, 3\pi/2$$

These are the more common  $90^\circ$  degree phase shifts.

We can generate a new phase table for the overall sequence. These phases,  $\phi_c$ , are to be applied to *all* of the pulses and receiver simultaneously.

$$\phi_c = 0 \ \pi/2 \ \pi \ 3\pi/2$$

$$\psi = 0 \ \pi/2 \ \pi \ 3\pi/2$$

This phase cycle is added “on top” of the other phase cycle to give the final table consisting of 16 steps.

$$\phi_1 = 2(0 \ \pi) \ 2(\pi/2 \ 3\pi/2) \ 2(\pi \ 0) \ 2(3\pi/2 \ \pi/2)$$

$$\phi_2 = 2(0) \ 2(\pi) \ 2(\pi/2) \ 2(3\pi/2) \ 2(\pi) \ 2(0) \ 2(3\pi/2) \ 2(\pi/2)$$

$$\psi = 2(0 \ \pi) \ 2(\pi/2 \ 3\pi/2) \ 2(\pi \ 0) \ 2(3\pi/2 \ \pi/2)$$

#### 6.4. Variants of COSY

Coherence transfer in COSY occurs through an antiphase state at state  $\langle B \rangle$  (Fig. 6.1.1). If the  $\langle A, B \rangle$  evolution period is short compared to  $1/(2J)$  then very little coherence is transferred and the resulting cross peaks are weak. In experiments that are collected rapidly, the limiting time is the number of points that are collected along  $t_1$ . Obviously, the number of  $t_1$  points must be kept small in order to keep the experimental time short. In the original COSY experiment, this would yield very little antiphase coherence at state  $\langle B \rangle$ , and therefore, very weak cross peaks would result. An alternative is to use a constant time for the  $t_1$  period [25]. This consists of the  $\langle A, B \rangle$  period being set to a fixed time, e.g.  $1/(2J)$ , and a  $180^\circ$  pulse systematically moved within this period to generate the  $t_1$  evolution period. When the  $180^\circ$  pulse is centered, there is no chemical shift evolution; however, since the  $180^\circ$  pulse is non-selective, coupling will proceed throughout the entire constant time period (Table 2.1B). As the  $180^\circ$  pulse is moved from the center there will be an increasing amount of chemical shift precession and frequency labeling of the coherence. In the constant time experiment, the amount of antiphase coherence at  $\langle B \rangle$  is constant. Since there is no time dependent modulation of the  $I$  coherence by coupling, the peaks along  $\omega_1$  will be singlets (decoupled from all other spins).

In situations where the transverse decay rate is fast compared to the reciprocal of the coupling constant, i.e. broad lines, the usable  $t_1$  period is limited, and therefore, the cross peaks are weak. Line broadening can arise from a variety of sources including chemical exchange or lifetime

broadening, paramagnetic broadening, or strong dipolar interactions due to long correlation times. With infinite sensitivity this does not cause a problem, but realistically, once the line width becomes a few times larger than the coupling constant the cross peaks become vanishingly small. The linewidth of protons in macromolecules becomes a limiting factor for COSY experiments at about 20,000 daltons. Above this molecular weight, the cross peaks are difficult to observe at typical concentrations of 1–10 mM.

## 7. MULTIPLE QUANTUM FILTERED SPECTROSCOPY

Multiple quantum filtration of correlation spectra allows the scientist to use the topologies of the scalar coupling network to unravel the complexities of the spin system under study. Multiple quantum effects were known long before the introduction of 2D spectroscopy, but the production and interpretation of these effects were difficult. Using hard pulses on non-equilibrium states of the density matrix, in general, produces coherences that can have more than one transverse component. These coherences evolve under a Hamiltonian that has contributions from all spins involved in the coherence. It is this property that can be exploited for the study of complex spin systems.

### 7.1. *Double quantum filtered COSY (DQFCOSY)*

Double quantum filtered COSY reduces the problem of the dispersive diagonal peak that exists in COSY spectra [26]. Figure 7.1.1 shows a DQFCOSY pulse sequence and its associated CFN. The spins are allowed to come to thermal equilibrium or a steady state at  $\langle A \rangle$ . The first  $90^\circ$  pulse generates transverse magnetization. During  $\langle A, B \rangle$  the spin precesses at its chemical shift frequency and simultaneously evolves via coupling into an antiphase state at  $\langle B \rangle$ . The second  $90^\circ$  pulse generates, among others, a double quantum coherence designated by the simultaneous transverse lines  $\langle B, D \rangle$  and  $\langle C, E \rangle$  for the **I** and **S** spins. The time period for  $\langle B, D \rangle$  (and  $\langle C, E \rangle$ ) is kept short to eliminate any chemical shift precession at the double quantum frequency. The final pulse recovers antiphase coherence from the double quantum state at  $\langle E \rangle$ . The  $I_z S^\pm$  antiphase coherence precesses at the chemical shift frequency of **S** and refocuses into observable, inphase magnetization during the acquisition time,  $\langle E, F \rangle$ .

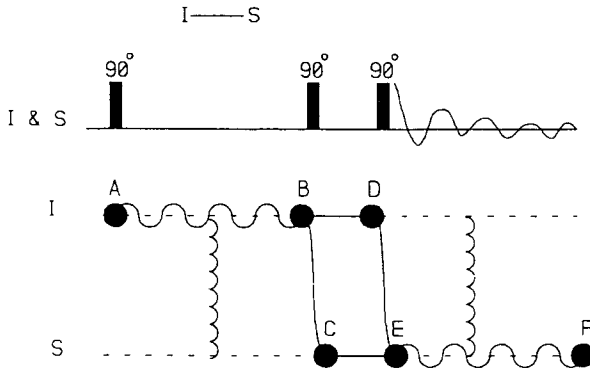


Fig. 7.1.1. CFN for double quantum filtered COSY.

The main difference between this sequence and the COSY sequence is that after the second  $90^\circ$  pulse in DQFCOSY the double quantum coherence is selected whereas in COSY the single quantum coherence is selected and detected. The double quantum coherence at  $\langle B, D \rangle \langle C, E \rangle$  is not directly observable and the final pulse is used to regenerate single quantum coherence. The selection of the double or single quantum coherence is accomplished by the particular phase cycle of the pulse sequence.

## 7.2. Product operator description of the DQFCOSY

*Preparation:*

$$\hat{I}_x \xrightarrow{\pi/2} -\hat{I}_y \quad (7.2.1)$$

*Evolution due to chemical shift:*

$$-\hat{I}_y \xrightarrow{\omega_I t_1} -\hat{I}_y \cos \omega_I t_1 + \hat{I}_x \sin \omega_I t_1 \quad (7.2.2)$$

*Evolution due to coupling:*

$$\begin{aligned} \xrightarrow{\pi J_{IS} t_1} & (-\hat{I}_y \cos \pi J_{IS} t_1 + 2\hat{I}_x \hat{S}_z \sin \pi J_{IS} t_1) \cos \omega_I t_1 \\ & + (\hat{I}_x \cos \pi J_{IS} t_1 + 2\hat{I}_y \hat{S}_z \sin \pi J_{IS} t_1) \sin \omega_I t_1 \end{aligned} \quad (7.2.3)$$

*First mixing pulse:*

$$\begin{aligned} \xrightarrow{\pi/2} & (-\hat{I}_x \cos \pi J_{IS} t_1 - 2\hat{I}_x \hat{S}_y \sin \pi J_{IS} t_1) \cos \omega_I t_1 \\ & + (\hat{I}_x \cos \pi J_{IS} t_1 - 2\hat{I}_y \hat{S}_y \sin \pi J_{IS} t_1) \sin \omega_I t_1 \end{aligned} \quad (7.2.4)$$



We can simplify the analysis at this point by applying a double quantum filter. A rudimentary double quantum filter can be implemented by subtracting this result (Eq. 7.2.4) from a result where the first two pulses are shifted in phase by  $90^\circ$ . The phase shift substitutes  $2\mathbf{I}_y\mathbf{S}_x$  for the  $2\mathbf{I}_x\mathbf{S}_y$  term in Eq. (7.2.4). The action of the filter is to retain only the term  $2\mathbf{I}_x\mathbf{S}_y$  in the first phase cycle step and the  $2\mathbf{I}_y\mathbf{S}_x$  term from the second step of the phase cycle. Individually, the  $2\mathbf{I}_x\mathbf{S}_y$  and  $2\mathbf{I}_y\mathbf{S}_x$  terms are superpositions of zero and double quantum coherence. This can be seen from the definitions of the raising and lowering operators (Section 3.2)

$$\mathbf{I}_x = 1/2(\mathbf{I}^+ + \mathbf{I}^-) \text{ and } \mathbf{I}_y = -i/2 (\mathbf{I}^+ - \mathbf{I}^-) \quad (7.2.5)$$

The two-spin order represented in Cartesian coordinates can be expanded to

$$2\mathbf{I}_x\mathbf{S}_y = -i/2 [(\mathbf{I}^+ + \mathbf{I}^-)(\mathbf{S}^+ - \mathbf{S}^-)] = -i/2[\mathbf{I}^+\mathbf{S}^+ - \mathbf{I}^+\mathbf{S}^- + \mathbf{I}^-\mathbf{S}^+ - \mathbf{I}^-\mathbf{S}^-]. \quad (7.2.6)$$

The operators with two raising or two lowering operators are double quantum states. The other terms in Eq. (7.2.6) that have one raising and one lowering operator are zero quantum coherences. A double quantum filter extracts the  $\mathbf{I}^+\mathbf{S}^+$  and the  $\mathbf{I}^-\mathbf{S}^-$  terms. Pure double quantum coherence is obtained by combination of  $2\mathbf{I}_x\mathbf{S}_y$  and  $2\mathbf{I}_y\mathbf{S}_x$ :

$$1/2i (\mathbf{I}^+\mathbf{S}^+ - \mathbf{I}^-\mathbf{S}^-) = (\mathbf{I}_x + i\mathbf{I}_y)(\mathbf{S}_x + i\mathbf{S}_y) - (\mathbf{I}_x - i\mathbf{I}_y)(\mathbf{S}_x - i\mathbf{S}_y), \quad (7.2.7)$$

which is equal to

$$1/2(2\mathbf{I}_x\mathbf{S}_y + 2\mathbf{I}_y\mathbf{S}_x) \quad (7.2.8)$$

The final  $90^\circ$  pulse of the mixing period, converts the double quantum state into antiphase coherence that evolves into observable magnetization during the detection period.

$$-(2\mathbf{I}_x\mathbf{S}_y + 2\mathbf{I}_y\mathbf{S}_x) \sin \pi J_{\text{IS}} t_1 \cos \omega_I t_1 = \pi/2 (\hat{\mathbf{I}}_x + \hat{\mathbf{S}}_x) \Rightarrow (2\mathbf{I}_x\mathbf{S}_z + 2\mathbf{I}_z\mathbf{S}_x) \sin \pi J_{\text{IS}} t_1 \cos \omega_I t_1 \quad (7.2.9)$$

The  $2\mathbf{I}_x\mathbf{S}_z$  term gives rise to a diagonal peak and the  $2\mathbf{I}_z\mathbf{S}_x$  gives rise to a cross peak. Since the cross-peak signal is only collected on every other scan, the signal strength in DQFCOSY is inherently a factor of two

lower than that of COSY for a spectrum collected in the same amount of time. The advantage of the DQFCOSY spectrum is that the diagonal and cross peaks are both antiphase and absorptive (see Fig. 6.2.1). Any cancellation that occurs in the cross peak due to a large linewidth will also affect the diagonal peak in the same manner. The detected signals for the diagonal peak will be:

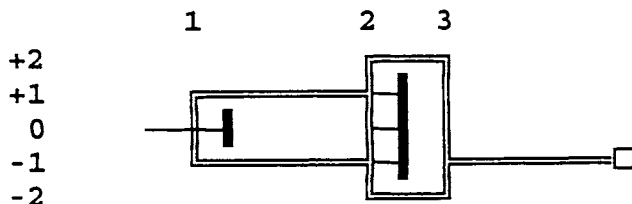
$$I_y \exp(i\omega_1 t_2) \sin \pi J_{IS} t_2 \sin \pi J_{IS} t_1 \cos \omega_1 t_1 \quad (7.2.10)$$

and for the cross peak:

$$S_y \exp(i\omega_S t_2) \sin \pi J_{IS} t_2 \sin \pi J_{IS} t_1 \cos \omega_1 t_1. \quad (7.2.11)$$

### 7.3. DQFCOSY phase cycling

The CTP for the DQFCOSY experiment is



With  $\Delta m_1 = \pm 1$ ,  $\Delta m_2 = 1, -3$ , and  $\Delta m_3 = -3, 1$  we obtain:

$\phi_1:$	$\phi_2$	$\phi_3$
-1 (0 1)	(-3 -2 -1 0) 1 2 3	(-3 -2 -1 0) 1 2 3
N=2	N=4	N=4
$\phi_1 = 0, \pi$	$\phi_2 = 0, \pi/2, \pi, 3\pi/2$	$\phi_3 = 0, \pi/2, \pi, 3\pi/2$

Phase table for DQFCOSY:

$\phi_1 = 4(0 \pi)$
$\phi_2 = 2(0) 2(\pi/2) 2(\pi) 2(3\pi/2)$
$\phi_3 = 8(0) 8(\pi/2) 8(\pi) 8(3\pi/2)$
$\psi = 0 \pi 3\pi/2 \pi/2 \pi 0 \pi/2 3\pi/2 3\pi/2 \pi/2 \pi 0 \pi/2 3\pi/2 0 \pi$
$\pi 0 \pi/2 3\pi/2 0 \pi 3\pi/2 \pi/2 \pi/2 3\pi/2 0 \pi 3\pi/2 \pi/2 \pi 0$

## 7.2. Variants

Multiple quantum filtering of COSY spectra is a wide ranging subject. The highest possible coherence order that can be generated is equal to the number of coupled spins in the system. Filters involving higher order coherences ( $>2$ ) can be used for selecting spin systems on the basis of number of spins [27] or even on the basis of the topology of the spin system [28].

One to the most used modifications of multiple quantum filtered COSY is E. COSY (Exclusive COSY) [29] or its variant P.E. COSY [30]. In E. COSY the cross peak patterns are modified by combining properly several orders of multiple quantum filtered COSY spectra. The simplification in the cross peak patterns allows for the accurate measurement of homonuclear coupling constants.

## 8. ISOTROPIC HOMONUCLEAR COHERENCE TRANSFER

### 8.1. Total correlation spectroscopy (TOCSY)

Total correlation spectroscopy (TOCSY) [31], also known as Homonuclear Hartmann–Hahn spectroscopy (HOHAHA) [32], has emerged in recent years to address the problem of correlating the chemical shifts of all of the spins in a scalar coupled network. The COSY experiment transfers coherence solely between spins that have a non-zero coupling constant. In a spin system consisting of several nuclei, there may be spins that have common coupling partners but have a zero or very small coupling between them. Several modifications of COSY, e.g. relayed COSY and double-relayed COSY, transfer coherence from one spin to a non-coupled spin via a coupled intermediate spin. However, these experiments have characteristics that make them rather unsuitable for the correlation of spins in large molecules. The TOCSY experiment uses a mixing period consisting of a series of phase modulated  $180^\circ$  pulses that “trick” all of the spins into precessing at the same frequency even though each spin has a unique chemical shift. By forcing the trajectory of all spins to be cyclic within the same time period the apparent energy levels for the resonances become identical. The ultimate sequence of RF pulses would create a state of *isotropic mixing* in which energy would flow unimpeded between all of the spins that are directly or indirectly coupled. Spins that have a zero (or very small) coupling to the peaks in

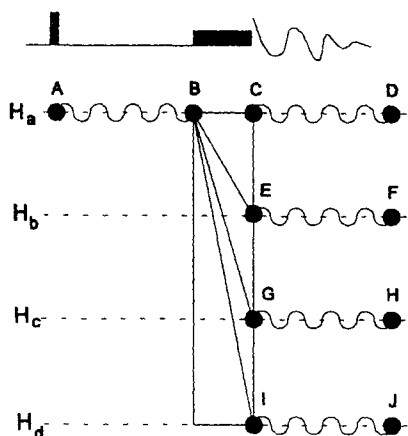


Fig. 8.1.1. Pulse sequence and CFN for total correlation spectroscopy (TOCSY).

the spin system do not join into this orgy of energy transfer. The pulse sequence and CFN for a TOCSY experiment are shown in Fig. 8.1.1.

The initial  $90^\circ$  pulse generates transverse magnetization at state  $\langle A \rangle$ . During  $\langle A, B \rangle$ , the I spin is frequency labeled with its chemical shift. Coupling between I and other spins also occurs during  $\langle A, B \rangle$ ; however, since inphase coherence is transferred during the mixing period, there is no essential active coupling that occurs in this period. The active coupling that does occur during  $\langle A, B \rangle$  is a source of phase distortions in the cross peaks. Many sequences contain a spin-lock-trim pulse prior to the mixing period that dephases components that are orthogonal to the spin lock axis. Spin-lock-trim pulses are also used at the end of the mixing period for the same purpose. Care must be taken so that the second spin lock does not rephase components that were dephased by the first spin lock. Usually, setting different lengths for the spin lock pulses avoids this problem.

An isotropic mixing sequence is applied during  $\langle B, C \rangle$  that transfers coherence among all of the spins in the coupled system. The first mixing sequences that were used were borrowed from techniques for heteronuclear decoupling, such as, WALTZ and MLEV. Heteronuclear decoupling requires that the spins rotate in a cyclic manner independent of resonance offset. The splitting of the heteronucleus due to coupling is eliminated when the decoupled spins are averaged to zero in a time frame that is small compared to the reciprocal of the heteronuclear coupling constant. The heteronuclear decoupling sequences are very

good at forcing the spins into offset-independent cyclic trajectories; however, they were designed without the consideration of homonuclear coupling interactions. The TOCSY experiment relies on homonuclear coupling and the sequences do not perform as well as one might think. Newer sequences with better performance, such as DIPSI [33] and FLOPSY [34], have been designed as cyclic decoupling sequences in the presence of homonuclear coupling.

After the isotropic mixing sequence, the magnetization that started on the **I** spin is distributed throughout the spin system. Two-dimensional Fourier transformation of the time domain data yields a spectrum with cross peaks among all of the spin system components. The product operator description of the TOCSY experiment is not as simple as most other experiments, in that during the isotropic mixing period the evolution of the spin system is similar to the evolution in a strongly coupled spin system. The basis functions that comprise the normal product operator formalism are for weakly coupled systems. The main complication is that the evolution of the spins under the coupling operator no longer commutes for separate couplings involving a given spin. For example, if a spin **I** is coupled to both **S** and **T**, then during the mixing period the evolution of the **I** spin depends *simultaneously* on both  $J_{IS}$  and  $J_{IT}$ ; they cannot be calculated independently. Modifications of the basis functions used in the product operator formalism can be made to correct for this problem; the methods are available in the literature [35].

The features of TOCSY transfer can be described in a simple manner if a two-spin system is considered [31]. The evolution of coherence has the form:

$$\begin{aligned} \mathbf{I}_x = 2\pi J_{IS}t \mathbf{I} \cdot \mathbf{S} \Rightarrow & 1/2 \mathbf{I}_x \{1 + \cos(2\pi J_{IS}t)\} + 1/2 \mathbf{S}_x \{1 + \cos(2\pi J_{IS}t)\} \\ & + (2\mathbf{I}_y \mathbf{S}_z - 2\mathbf{I}_z \mathbf{S}_y) \sin(2\pi J_{IS}t) \end{aligned} \quad (8.1.1)$$

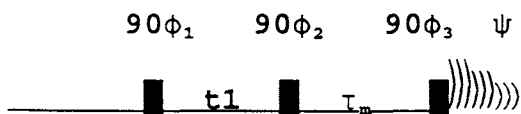
where  $\mathbf{I} \cdot \mathbf{S}$  represents the isotropic coupling operator. Similar expressions arise for  $\mathbf{I}_y$  and  $\mathbf{I}_z$  components. Since the mixing is (ideally) isotropic, it occurs for transverse or longitudinal components. Antiphase components also are transferred. The noteworthy point for this type of transfer is that inphase  $\mathbf{I}_x$  is transferred directly to inphase  $\mathbf{S}_x$ . The resulting cross peaks are, therefore, also inphase. Although the transfer appears to “prevent the canceling” that occurs in antiphase transfer involving broad lines, the coherence transfer still relies on the coupling constant. If the spectral lines are significantly broader than the coupling constant the transfer will be quenched.

## 9. INCOHERENT MAGNETIZATION TRANSFER

### 9.1. Nuclear Overhauser effect spectroscopy (NOESY)

The nuclear Overhauser effect (NOE) is the feature of NMR spectroscopy that provides for the determination of the spatial distance between two nuclei. It is this property that is the major contributor to the determination of the three-dimensional structure of macromolecules in solution. Much effort has been placed in the methods to extract accurate distance information from NOESY spectra and the use of these data in the determination of three dimensional structures. A number of books and reviews cover the theory and experimental details of modern magnetization exchange spectroscopy, including NOESY, EXSY and ROESY (Rotating frame Overhauser Effect Spectroscopy) [36–38]. Here I will discuss only the basic NOESY experiment.

The CFN for NOESY is identical to that of the EXSY (Fig. 4.1.3) experiment described earlier. The transfer of magnetization by the NOE is formally identical to magnetization exchange by physical movement of the nuclei. The pulse sequence and the phase cycling are also identical between NOESY and EXSY. In a normal NOESY spectrum the exchange cross peaks are indistinguishable from the NOE cross peaks. The pulse sequence and the product operator description of NOESY are as follows:



*Preparation:*

$$\mathbf{I}_z + \mathbf{S}_z = \pi/2 \hat{\mathbf{I}}_x \Rightarrow -\mathbf{I}_y - \mathbf{S}_y \quad (9.1.1)$$

*Evolution:*

$$\begin{aligned} = \omega_I t \hat{\mathbf{I}}_z + \omega_S t \hat{\mathbf{S}}_z \Rightarrow & \quad -\mathbf{I}_y \cos \omega_I t_1 + \mathbf{I}_x \sin \omega_I t_1 \\ & -\mathbf{S}_y \cos \omega_S t_1 + \mathbf{S}_x \sin \omega_S t_1 \end{aligned} \quad (9.1.2)$$

*Mixing:*

$$\begin{aligned} = \pi/2 (\hat{\mathbf{I}}_x + \hat{\mathbf{S}}_x) \Rightarrow & \quad -\mathbf{I}_z \cos \omega_I t_1 + \mathbf{I}_x \sin \omega_I t_1 \\ & -\mathbf{S}_z \cos \omega_S t_1 + \mathbf{S}_x \sin \omega_S t_1 \end{aligned} \quad (9.1.3)$$

Applying a *zero* order coherence filter by phase cycling (Section 4.3), the state becomes

$$-\mathbf{I}_z \cos \omega_I t_1 - \mathbf{S}_z \cos \omega_S t_1. \quad (9.1.4)$$

The system now undergoes cross relaxation according to the master equation.

$$\mathbf{M}(t) = \exp(-\mathbf{R}\tau_m)\Delta\mathbf{M}(0) \quad (9.1.5)$$

where  $\Delta\mathbf{M}(0)$  is vector containing the deviations of the magnetization from equilibrium immediately after the first  $90^\circ$  pulse of the mixing period and  $\mathbf{M}(t)$  is the vector of magnetization after a time,  $\tau_m$ , of cross relaxation. The relaxation matrix,  $\mathbf{R}$ , contains the auto- and cross-relaxation rate constants. During  $\tau_m$  magnetization “exchanges” between the  $\mathbf{I}$  and  $\mathbf{S}$  nuclei. Without concerning ourselves with the details, the state at the end of the mixing period,  $\tau_m$ , can be simply represented as

$$-\mathbf{I}_z (A \cos \omega_I t_1 + B \cos \omega_S t) - \mathbf{S}_z (C \cos \omega_S t_1 + D \cos \omega_I t) \quad (9.1.6)$$

where the coefficients A through D represent amplitudes due to auto- and cross-relaxation of the spin system. In effect, some  $\mathbf{I}$  magnetization has become  $\mathbf{S}$  magnetization and *vice versa*. To end the mixing period a  $90^\circ$  pulse is applied.

$$\begin{aligned} = \pi/2(\hat{\mathbf{I}}_x + \hat{\mathbf{S}}_x) \Rightarrow & \mathbf{I}_y (A \cos \omega_I t_1 + B \cos \omega_S t) \\ & + \mathbf{S}_y (C \cos \omega_S t_1 + D \cos \omega_I t) \end{aligned} \quad (9.1.7)$$

*Detection:*

$$\begin{aligned} = \omega_I t \hat{\mathbf{I}}_z + \omega_S t \hat{\mathbf{S}}_z \Rightarrow & \mathbf{I}_y \exp(i\omega_I t_2) (A \cos \omega_I t_1 + B \cos \omega_S t) \\ & + \mathbf{S}_y \exp(i\omega_S t_2) (C \cos \omega_S t_1 + D \cos \omega_I t) \end{aligned} \quad (9.1.8)$$

The detected signals, upon two-dimensional Fourier transform, yield a map of all of the spins that are close in space.

## 10. HETERONUCLEAR SHIFT CORRELATIONS

The correlation of the chemical shifts of heteronuclei is usually accomplished by the detection of the more sensitive proton spin. Since the time when this technique was introduced by Mueller [39], a number of variants have been developed [40–44,46].





during <C,D>. The second  $90^\circ$  **S** pulse at <D,E> recreates the  $2\mathbf{I_xS_z}$  antiphase state at <E> that is now frequency labeled with the **S** chemical shift. The <E,F> period serves to refocus the antiphase state into  $\mathbf{I_y}$  at <F> which is then detected during <F,G> with **S** nucleus decoupling.

*Preparation:*

$$\mathbf{I_z} = \pi/2 \hat{\mathbf{I_x}} \Rightarrow -\mathbf{I_y} \quad (10.1.1)$$

$$-\mathbf{I_y} = \pi J_{\text{IS}} t 2\hat{\mathbf{I_x}}\hat{\mathbf{S_z}} \Rightarrow -\mathbf{I_y} \cos \pi J_{\text{IS}} t + 2\mathbf{I_xS_z} \sin \pi J_{\text{IS}} t \quad (10.1.2)$$

With  $t = 1/(2J_{\text{IS}})$ ,  $\cos \pi J_{\text{IS}} t = 0$  and  $\sin \pi J_{\text{IS}} t = 1$ , the state at <B> (Eq. 10.1.2) becomes,

$$2\mathbf{I_xS_z} \quad (10.1.3)$$

*Evolution:*

The  $\pi/2$  pulse on the **S** spins yields a coherence in which both spins are transverse. This is a superposition of zero and double quantum coherence (Section 7.2.6).

$$2\mathbf{I_xS_z} = \pi/2 \hat{\mathbf{S_x}} \Rightarrow -2\mathbf{I_xS_y} \quad (10.1.4)$$

This coherence evolves only under the chemical shift of the **S** spin due to the refocusing effect of the  $180^\circ$  pulse on the **I** spin (Eqs. 10.1.5–10.1.7)

$$-2\mathbf{I_xS_y} = \omega_{\text{St}_1}/2 \hat{\mathbf{S_z}} \Rightarrow -2\mathbf{I_xS_y} \cos \omega_{\text{St}_1}/2 + 2\mathbf{I_xS_x} \sin \omega_{\text{St}_1}/2 \quad (10.1.5)$$

$$= \pi \hat{\mathbf{I_x}} \Rightarrow -2\mathbf{I_xS_y} \cos \omega_{\text{St}_1}/2 + 2\mathbf{I_xS_x} \sin \omega_{\text{St}_1}/2 \quad (10.1.6)$$

$$= \omega_{\text{St}_1}/2 \hat{\mathbf{S_z}} \Rightarrow -2\mathbf{I_xS_y} \cos \omega_{\text{St}_1} + 2\mathbf{I_xS_x} \sin \omega_{\text{St}_1}. \quad (10.1.7)$$

The second **S**  $\pi/2$  pulse recreates antiphase coherence labeled by the frequency of the **S** spin

$$= \pi/2 \hat{\mathbf{S_x}} \Rightarrow -2\mathbf{I_xS_z} \cos \omega_{\text{St}_1} + 2\mathbf{I_xS_x} \sin \omega_{\text{St}_1}. \quad (10.1.8)$$

$2\mathbf{I_xS_x}$  is unobservable and can be neglected. Further evolution under coupling gives,

$$= \pi J_{IS} t \hat{2} \hat{I}_z \hat{S}_z \Rightarrow (-2 \mathbf{I}_x \mathbf{S}_z \cos \pi J_{IS} t - \mathbf{I}_y \sin \pi J_{IS} t) \cos \omega_{St_1} \quad (10.1.9)$$

Again with  $t = 1/(2J_{IS})$ , one obtains

$$= -\mathbf{I}_y \cos \omega_{St_1}. \quad (10.1.10)$$

*Detection:*

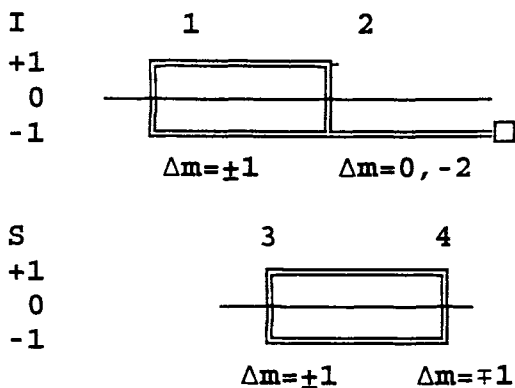
With **S** nucleus decoupling, the only evolution during detection is due to **I** chemical shift.

$$-\mathbf{I}_y \cos \omega_{St_1} \exp(i \omega_1 t_2) \quad (10.1.11)$$

With quadrature detection in  $\omega_1$  and hypercomplex Fourier transform, the two-dimensional spectrum consists of peaks located at the intersection of the chemical shifts of the scalar coupled **I** and **S** spins (as in Fig. 5.1.4).

## 10.2. HMQC phase cycling

For heteronuclei, nuclear-spin selective pulses do not effect the evolution of the heteronucleus. For this reason, the phase cycling is done independently.



Performing the necessary phase calculations:

<b>I</b> $\phi_1$ :	<b>I</b> $\phi_2$ :	<b>S</b> $\phi_3$ :	<b>S</b> $\phi_4$ :
$\Delta m = \pm 1$	$\Delta m = 0, -2$	$\Delta m = \pm 1$	$\Delta m = \mp 1$
+1 (0 -1)	(-2 -1) 0 1 2	-1 (0 +1)	-1 (0 +1)
$N=2$	$N=2$	$N=2$	$N=2$
$\phi_1 = 0 \pi$	$\phi_2 = 0 \pi$	$\phi_3 = 0 \pi$	$\phi_4 = 0 \pi$

To remove quadrature images, the **I** channel can be further constrained to allow only  $\Delta m = -1$  for the entire sequence. This phase cycle is the CYCLOPS procedure of Section 3.9. With the CYCLOPS sequence superimposed on the rest of the **I** channel phase cycle gives a 64 step phase cycle. Again, note that the **S** channel is treated independently.

64 step phase table for HMQC:

$$\begin{aligned}
 \phi_1: & 8(0, \pi) \ 8(\pi/2, 3\pi/2) \ 8(\pi, 0) \ 8(3\pi/2, \pi/2) \\
 \phi_2: & 4(0, 0, \pi, \pi) \ 4(\pi/2, \pi/2, 3\pi/2, 3\pi/2) \ 4(\pi, \pi, 0, 0) \ 4(3\pi/2, 3\pi/2, \pi/2, \pi/2) \\
 \phi_3: & 8(0, 0, 0, 0, \pi, \pi, \pi, \pi) \\
 \phi_4: & 4(0, 0, 0, 0, 0, 0, \pi, \pi, \pi, \pi, \pi, \pi) \\
 \psi: & (0, \pi, \pi, 0, \pi, 0, 0, \pi, \pi, 0, 0, \pi, 0, \pi, \pi, 0) \\
 & (3\pi/2, \pi/2, \pi/2, 3\pi/2, \pi/2, 3\pi/2, 3\pi/2, \pi/2, \pi/2, 3\pi/2, \pi/2, 3\pi/2, \pi/2, \pi/2, 3\pi/2) \\
 & (\pi, 0, 0, \pi, 0, \pi, \pi, 0, 0, \pi, \pi, 0, \pi, 0, 0, \pi) \\
 & (\pi/2, 3\pi/2, 3\pi/2, \pi/2, 3\pi/2, \pi/2, \pi/2, 3\pi/2, 3\pi/2, \pi/2, \pi/2, 3\pi/2, \pi/2, 3\pi/2, 3\pi/2, \pi/2)
 \end{aligned}$$

This represents a complete phase cycle for the HMQC experiment. One major disadvantage is the number of experiments necessary to accomplish the full phase cycle. A modest size for a two-dimensional experiment would be to collect 512 points along  $t_1$ . If the recycle time for each scan is 1 s then the minimum time for an HMQC experiment would be  $1 \times 64 \times 512 = 32768$  seconds or about 9 hours. The sensitivity of this experiment for a fully  $^{13}\text{C}$  or  $^{15}\text{N}$  labeled sample (the **S** nucleus) is very nearly the sensitivity obtained for  $^1\text{H}$  (the **I** spin) alone. A sample of 1–2 mM concentration gives good signal-to-noise for  $^1\text{H}$  spectra at 500 MHz in a very short period of time (<5 minutes). Ideally, the 2D experiment should only take a few minutes, but with full phase cycling the time is enormously increased. An obvious solution is not to do the complete phase cycle. The first luxury to be eliminated is the CYCLOPS routine which eliminates quadrature images. If the spectrometer receiver is adjusted properly, the amplitude of the quadrature images should be less than 1% of the real signal. Eliminating this step of the phase cycle reduces the number of required scans by a factor of 4. This still means that the experiment will take a little more than 2 hours to collect. Heteronuclear experiments are unique in that the RF pulses applied to one spin do not effect the heteronuclear spin. The part of the phase cycle in which the **S** spin pulse phase is inverted along with the inversion of the receiver eliminates any **I** coherence that is not part of a heteronuclear spin coherence. This selects only **I** spins that are coupled to **S** spins

and eliminates all other uncoupled **I** spins, which are unmodulated by the heteronuclear coupling. This spectral editing feature is very effective as a phase cycling step. Applying the most severe reduction of phase cycling, each  $t_1$  block of the HMQC experiment can be obtained by 2 acquisitions.

$$\begin{aligned}\phi_3 & 0 \pi \\ \psi & 0 \pi\end{aligned}$$

This reduces the experimental time to about 20 minutes. More rapid accumulations can be made by reducing the resolution in the  $t_1$  dimension or reducing the recycle time.

The steps in a phase cycle that can be eliminated or truncated are not always obvious. As a rule of thumb, in a heteronuclear sequence, the two step phase cycle of the non-observed nucleus that inverts the coherence and thus inverts the receiver is the simplest sequence. Additional phase cycles on the observed or unobserved nucleus that are the next most efficient depend on the particular type of artifacts that are to be removed.

### 10.3. Modulation from $^1\text{H}$ - $^1\text{H}$ scalar coupling

In most cases there will be other spins (**I'**) in the molecule under study that are coupled to **I**. During the  $t_1$  period, **I**-**I'** coupling is active, and the **S** spin is modulated by  $\cos(\pi J_{\text{II}'} t_1)$  during the  $t_1$  period (Fig. 10.3.1). If the  $t_1$  period is kept short compared to the reciprocal of the small **I**-**I'** coupling constant, the passive coupling will not influence the **S** signal to any great extent. However, in order to increase the frequency resolution in the **S** dimension, the  $t_1$  period must be lengthened. The result of this passive coupling during  $t_1$  is that the peak is split along the  $\omega_S$  dimension by  $J_{\text{II}'}$ . This method has been used to obtain  $^3J_{\alpha\beta}$   $^1\text{H}$  coupling constants in  $^{15}\text{N}$  labeled proteins.

### 10.4. Heteronuclear single quantum correlation (HSQC)

Figure 10.4.1 shows the pulse sequence and CFN for Heteronuclear Single Quantum Correlation (HSQC) [43]. Like HMQC, this sequence is used for the chemical shift correlation of **I** and **S** nuclei with different magnetogyric ratios. The sequence consists of a pair of symmetric transfer steps that surround the chemical shift evolution period of the

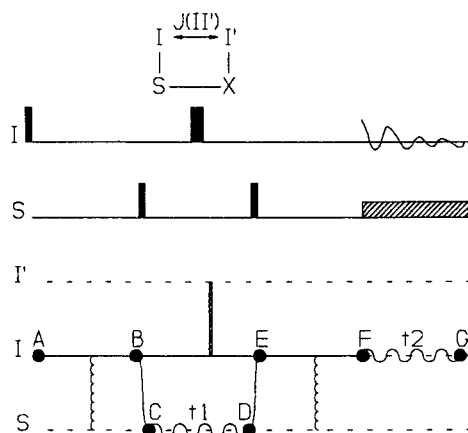


Fig. 10.3.1. HMQC pulse sequence with CFN showing passive  $I'$  coupling.

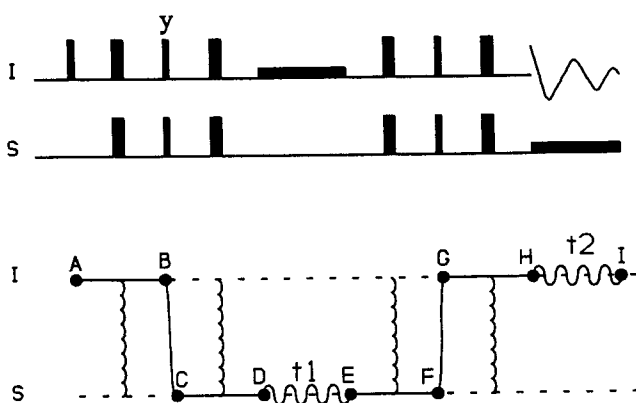


Fig. 10.4.1. Pulse sequence and CFN for heteronuclear single quantum spectroscopy. This particular experiment is a double refocused INEPT type transfer.

S nucleus. After the first  $90^\circ$  pulse on the I spin at <A>, the system evolves by heteronuclear scalar coupling to a heteronuclear antiphase state,  $2I_xS_z$ , at <B>. Coherence transfer occurs along <B,C> and at <C> the state of the system is  $2I_xS_y$ . The antiphase state at <C> is refocused into pure  $-S_x$  coherence during <C,D>. Frequency labeling of the S spin occurs during the  $t_1$  period <D,E>. The IS scalar coupling interaction is

removed during  $\langle D, E \rangle$  by continuous broadband spin decoupling of the  $I$  spin. The decoupling can be accomplished also by replacing the broadband decoupling sequence with a single  $180^\circ$  pulse at the center of the  $t_1$  period. The resulting  $S$  resonance is slightly narrower in the sequence using broadband decoupling. From the end of the  $t_1$  period, the sequence consists of a mirror image of the  $\langle A, D \rangle$  path. Antiphase coherence is obtained during  $\langle E, F \rangle$ , coherence transfer occurs across  $\langle F, G \rangle$ , and the resulting antiphase state is refocused into pure  $I$  spin coherence during  $\langle G, H \rangle$  for detection during  $t_2$  ( $\langle H, I \rangle$ ).

Using product operator analysis, the first pulse generates

$$I_z = \pi/2 \hat{I}_x \Rightarrow -I_y \quad (10.4.1)$$

The sequence of a delay followed by a  $180^\circ$  pulse on the  $x$  axis (phase equal to zero) on both coupled nuclei followed by the same delay can be represented as a single rotation (Table 2.1B).

$$= \pi J_{IS} \Delta 2 \hat{I}_z \hat{S}_z \Rightarrow \quad (10.4.2)$$

If  $\Delta$  is  $1/(2J_{IS})$ , then the sequence  $\langle A, B \rangle$  becomes:

$$I_z = \pi/2 \hat{I}_x \Rightarrow -I_y = \pi/2_{IS} 2 \hat{I}_z \hat{S}_z \Rightarrow 2 I_x S_z. \quad (10.4.3)$$

Simultaneous (or nearly so)  $90^\circ$  pulses on both nuclei effect the coherence transfer,  $\langle B, C \rangle$

$$2 I_x S_z = \pi/2 (\hat{I}_y + \hat{S}_x) \Rightarrow 2 I_x S_y. \quad (10.4.4)$$

Note that the phase of the  $I$  pulse at  $\langle B \rangle$  is  $90^\circ$  out of phase with respect to the first  $90^\circ$   $I$  pulse. The sequence  $\langle C, D \rangle$  can be simplified to

$$2 I_x S_y = \pi J_{IS} \Delta 2 \hat{I}_z \hat{S}_z \Rightarrow -S_x. \quad (10.4.5)$$

During the  $t_1$  period, the  $S$  coherence is frequency labeled,

$$-S_x = \omega_S t_1 \hat{S}_z \Rightarrow -S_x \cos \omega_S t_1 - S_y \sin \omega_S t_1 \quad (10.4.6)$$

During  $\langle E, F \rangle$ , the  $S$  transverse magnetization evolves into an antiphase state. Using the composite rotation from Eq. (10.4.2), we obtain

$$-\mathbf{S}_x \cos \omega_{st_1} - \mathbf{S}_y \sin \omega_{st_1} = \pi J_{IS} \Delta \hat{\mathbf{I}}_z \hat{\mathbf{S}}_z \Rightarrow -2\mathbf{I}_z \mathbf{S}_y \cos \omega_{st_1} + 2\mathbf{I}_z \mathbf{S}_x \sin \omega_{st_1}. \quad (10.4.7)$$

Coherence transfer occurs at <F,G> upon the application of 90° pulses to both spins.

$$-2\mathbf{I}_z \mathbf{S}_y \cos \omega_{st_1} + 2\mathbf{I}_z \mathbf{S}_x \sin \omega_{st_1} = \pi/2 (\hat{\mathbf{I}}_x + \hat{\mathbf{S}}_x) \Rightarrow 2\mathbf{I}_y \mathbf{S}_z \cos \omega_{st_1} - 2\mathbf{I}_y \mathbf{S}_x \sin \omega_{st_1} \quad (10.4.8)$$

Only the  $2\mathbf{I}_y \mathbf{S}_z$  term will evolve into observable magnetization since the  $2\mathbf{I}_y \mathbf{S}_x$  term is a superposition of heteronuclear multiple quantum states.

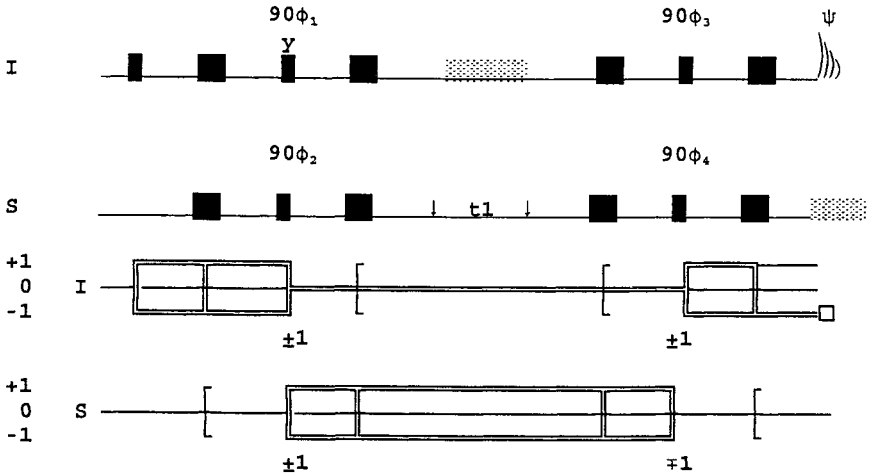
During <G,H> the antiphase state refocuses into observable **I** magnetization

$$2\mathbf{I}_y \mathbf{S}_z \cos \omega_{st_1} = \pi J_{IS} \Delta \hat{\mathbf{I}}_z \hat{\mathbf{S}}_z \Rightarrow -\mathbf{I}_x \cos \omega_{st_1}. \quad (10.4.9)$$

The detection period consisting of chemical shift on the **I** spins during <H,I> yields

$$-\mathbf{I}_x \cos \omega_{st_1} e^{i\omega(I)t} \quad (10.4.10)$$

### 10.5. Coherence transfer pathway and phase cycling for HSQC



A computation of the number of phase shifted experiments required for this sequence yields more than 8000 if all pulses are fully cycled. Obviously, this is not very practical. The pulses ( $\phi_1.. \phi_4$ ) that are marked in the above pulse sequence constitute editing steps. These steps involve the inversion of the coherence of interest followed by the subtraction of the resultant signal in the computer memory.

Consider the  $90^\circ$  pulse with phase  $\phi_2$  at the coherence transfer step  $\langle B, C \rangle$ . The standard analysis for the phase cycle yields:

$\phi_2$ :

$$\Delta m = \pm 1$$

$$(1\ 0) \rightarrow -1$$

$$N=2$$

$$\phi_2 = 0, \pi$$

For the coherence at  $\langle B \rangle$  (Fig. 10.4.1), applying a  $90^\circ S_x$  and a  $90^\circ I_y$  pulse yields

$$2I_x S_z = \pi/2 (\hat{I}_y + \hat{S}_x) \Rightarrow 2I_z S_y. \quad (10.5.1)$$

Eventually this leads to the detected signal

$$-I_x \cos \omega_s t_1. \quad (10.5.2)$$

Inverting the phase of the  $S$  pulse ( $-90^\circ S_x$ ) changes the sign of the coherence;

$$2I_x S_z = \pi/2 \hat{I}_y - \pi/2 \hat{S}_x \Rightarrow -2I_z S_y \quad (10.5.3)$$

leading to a detected signal of opposite sign.

$$I_x \cos \omega_s t_1 \quad (10.5.4)$$

In order to have constructive interference, the signals from these two steps in the phase cycle the two signal must be subtracted. Any other  $I$  coherence that is not in an antiphase state with  $S$  will not be inverted in either sequence and any resultant observable signal will be subtracted away. For example, if there is a spin  $I'$ , which is not coupled to an  $S$  spin, its behavior at step  $\langle B, C \rangle$  will be:

$$I'_y = \pi/2 (\hat{I}'_y + \hat{S}_x) \Rightarrow I'_y \quad (10.5.5)$$



and

$$\mathbf{I}'_y = \pi/2 \hat{\mathbf{I}}'_y - \pi/2 \hat{\mathbf{S}}_x \Rightarrow \mathbf{I}'_y \quad (10.5.6)$$

Any part of this  $\mathbf{I}'$  signal, which ultimately reaches the receiver, will be removed by subtraction.

The CTP analysis for all of the other  $90^\circ$  pulses is identical to this one since all of the changes on coherence order are  $\pm 1$  (or  $\mp 1$ ). The same effect is obtained by inverting the  $\mathbf{I}$  pulse at step  $\langle \mathbf{B}, \mathbf{C} \rangle$ .

$$2\mathbf{I}_x \mathbf{S}_z = -\pi/2 \hat{\mathbf{I}}_y + \pi/2 \hat{\mathbf{S}}_x \Rightarrow -2\mathbf{I}_x \mathbf{S}_y \quad (10.5.7)$$

and the resulting signal will be subtracted from the first experiment  $[\pi/2(\mathbf{I}_y + \mathbf{S}_x) = \lambda]$  and added to the result from the second phase cycle step  $[\pi/2(\mathbf{I}_y - \pi/2 \mathbf{S}_x) = \lambda]$ . The phase table for this experiment with only the  $90^\circ$  pulses cycled gives a phase cycle with a length of 16 scans.

$$\begin{aligned} \phi_1: & 2(0) \ 2(\pi) \\ \phi_2: & 8(0 \ \pi) \\ \phi_3: & 8(0) \ 8(\pi) \\ \phi_4: & 4(0) \ 4(\pi) \\ \psi: & (0 \ \pi \ \pi \ 0) \ 2(\pi \ 0 \ 0 \ \pi) \ (0 \ \pi \ \pi \ 0) \end{aligned}$$

All of the other pulses in this sequence, except for the initial  $90^\circ$  pulse, are  $180^\circ$  pulses. These pulses are rather sensitive to off-resonance effects. There are two different types of phase cycling schemes that are usually employed for  $180^\circ$  pulses depending on whether the magnetization is transverse or longitudinal. For  $180^\circ$  pulses used to invert  $Z$  magnetization, the phase can be cycled in  $90^\circ$  steps without changing the receiver phase (Section 3.11). For transverse magnetization EXORCYCLE is used; this phase cycle consists of  $90^\circ$  phase increments associated with phase inversion of the receiver (Section 3.10). To shorten the length of the experiment, often the phases of  $180^\circ$  pulses are simply inverted without any changing the receiver phase.

The quality of  $180^\circ$  degree pulses used to invert  $Z$  magnetization can be improved by using composite pulses of the type  $90_x:180_y:90_x$  or  $90_x:180_{-x}:270_x$ . This may lead to greater sensitivity in certain experiments. Special care must be taken in applying composite  $180^\circ$  pulses to transverse magnetization and this procedure is not generally recommended without detailed analysis.

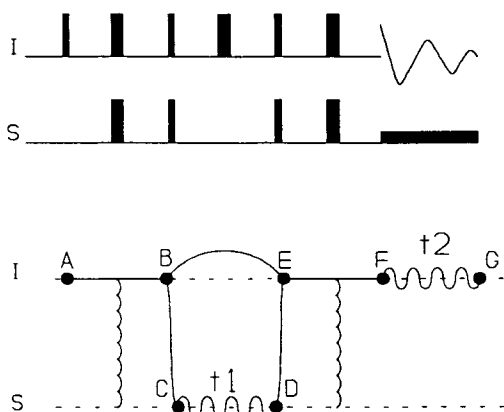


Fig. 10.6.1. Pulse sequence and CFN for a HSQC experiment with evolution of **S** antiphase with respect to **I**.

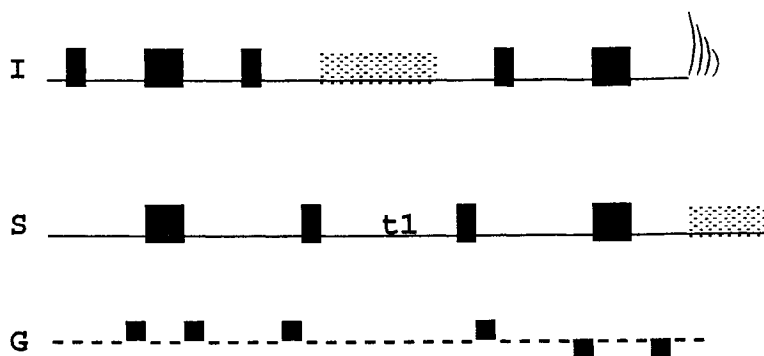
### 10.6. Variations on HSQC

The HSQC experiment can also be collected with the pulse sequence shown in Fig. 10.6.1. The main difference between this sequence and the experiment described in Fig. 10.4.1 is that the antiphase signal,  $2I_zS_y$ , at state  $\langle C \rangle$  is not refocused into  $S_x$ . The refocusing step is eliminated on both sides of the  $t_1$  period and the chemical shift evolution of the **S** spin occurs as an antiphase state during  $\langle C, D \rangle$ . The Z primitive (Table 1.1) that connects states  $\langle B \rangle$  and  $\langle E \rangle$  represents the retention of the antiphase state during this period. The  $180^\circ$  pulse in the center of  $\langle C, D \rangle$  is a decoupling pulse, which replaces the broadband decoupling sequence shown in Fig. 10.4.1. In detail, the use of a decoupling sequence, such as WALTZ or MLEV, prevents oscillation between antiphase and inphase states, whereas a single  $180^\circ$  pulse permits antiphase to inphase oscillation during  $\langle B, E \rangle$  with the antiphase state returned at  $\langle E \rangle$ . This difference becomes important if the relaxation of the of the antiphase state is of interest. The oscillation between antiphase and inphase states will allow transverse relaxation from both states. This will be manifested as a superposition of **S** resonances with different linewidths. In this sequence, if a broadband decoupling cycle is used then care must be taken to assure that the **I** spins are returned to the Z axis at  $\langle E \rangle$ .

Another variation of heteronuclear shift correlation experiments is heteronuclear single- and multiple-quantum correlation (HSMQC) [45]. HSMQC is a combination of HMQC and HSQC, which provides improved sensitivity by the elimination of coupling to remote protons. Essentially, this technique superimposes HMQC and HSQC cross peaks.

An enormous reduction in the time required to collect HSQC spectra can be achieved by the use of pulsed field gradients (PFG). The number of acquisitions required per  $t_1$  increment can be lowered to one. The HSQC experiment lends itself very nicely to the use of PFGs since there are a number of "sites" in the pulse sequence where they can be incorporated. The two especially vulnerable sites are at the points of heteronuclear coherence transfer. By the proper placement of the  $^1\text{H}$  and the X pulses, a longitudinal two-spin order state  $2\text{I}_z\text{S}_z$  can be created. The coherence order of this state is zero and thus has no dependence on a phase shift (Eq. 3.12.3). Since these points are editing steps of this sequence, only the desired magnetization passes the PFG spoilers. Another advantage of the use of PFGs is that since the diffusion coefficient for water is much larger than that for a macromolecule, the water resonance can be effectively eliminated without the use of presaturation. Any loss of intensity of amide proton signals due to saturation transfer from water and any intensity losses due to spin diffusion from saturated  $\alpha$  protons having the same chemical shift as water can be circumvented.

One possible placement of the PFGs in the HSQC experiment is shown below.



The PFGs surrounding the  $180^\circ$  pulses were discussed in Section 3.14.

## 11. CONCLUSION

Two-dimensional methods revolutionized the study of macromolecules by NMR spectroscopy. At the beginnings of macromolecular NMR, the only options available to the scientist were one-dimensional spectra of well resolved resonances. Any analysis of NMR experiments still requires well resolved resonances; however, the resolution toolkit from which the spectroscopist can draw is much larger now. The current trend in the study of larger molecules is to obtain 3D, 4D and even higher dimensional spectra of isotopically labeled molecules. The logic being that higher resolution is required, however, does not leave two-dimensional NMR in the dust. It is always a good idea to obtain data from the simplest possible experiment. If the data can be unambiguously obtained from an one-dimensional NMR spectrum, then there is no real need to use two-dimensional spectroscopy. Likewise, many problems associated with large molecular systems can be solved by simple two-dimensional methods. The scientific questions should rule the technology.

In the same light, if it is impossible to obtain the data with one dimensional data, then the increased complexity of two or higher dimensions should not stand as a barrier to their use. A major stumbling block of going to higher dimensional spectra is that most of the experiments are not readily explained by the standard classical vector models. The 'magic' of phase cycling also often causes great amounts of undue stress. In this review, I have not scratched the surface of the hundreds of two-dimensional experiments that are available from the literature. I have been lax in mentioning the most modern experiments for obtaining relaxation times and coupling parameters. I hope, however, that the visual approach of the CFN can be useful for the understanding of these powerful experiments. The top-down approach has been very useful in our laboratory for the transfer of NMR technology to the practising biochemist, as well as supplying a simple graphical language that can be used to discuss new experiments [46].

## REFERENCES

1. R.R. Ernst, G. Bodenhausen and A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Oxford University Press, New York, 1987.
2. J. Jeener, *Ampère International Summer School*. Basko Polje, Yugoslavia, 1971.

3. W.P. Aue, E. Bartholdi and R.R. Ernst, *J. Chem. Phys.*, 64 (1976) 2229.
4. M. Goldman, *Quantum Description of High-Resolution NMR in Liquids*. Oxford University Press, New York, 1988.
5. D.E. Wemmer, *Concepts in Magnetic Resonance*, 1 (1989) 59.
6. A. Bax and L. Lerner, *Science*, 232 (1986) 960.
7. H. Kessler, M. Gehrke and C. Griesinger, *Angew. Chem. Int. Ed. Engl.*, 27 (1988) 490.
8. O.W. Sørensen, G.W. Eich. M.H. Levitt, G. Bodenhausen and R. R. Ernst, *Prog. NMR Spectrosc.*, 16 (1983) 163.
9. R. Benn and H. Günther, *Angew. Chem. Int. Ed. Engl.*, 22 (1983) 350.
10. C.J. Turner, *Prog. NMR Spectrosc.*, 16, (1984) 311.
11. O.W. Sørensen, G.W. Eich. M.H. Levitt, G. Bodenhausen and R. R. Ernst, *Prog. NMR Spectrosc.*, 16 (1983) 163.
12. F.J.M. van de Ven and C.W. Hilbers, *J. Magn. Reson.*, 54 (1983) 512.
13. K.J. Packer and K.M. Wright, *Mol. Phys.*, 50 (1983) 797.
14. U. Eggenberger and G. Bodenhausen, *Angew. Chem. Int. Ed. Engl.*, 29 (1990) 374.
15. M. Ikura, L.E. Kay and A. Bax, *J. Magn. Reson.*, 29 (1990) 496.
16. G.M. Clore and A.M. Gronenborn, *Prog. NMR Spectrosc.*, 23 (1991) 43.
17. A.D. Bain, *J. Magn. Reson.*, 56 (1984) 418.
18. E.O. Stejskal and J. Schaefer, *J. Magn. Reson.*, 14 (1974) 160.
19. G. Bodenhausen, R. Freeman and D.L. Turner, *J. Magn. Reson.*, 27 (1977) 511.
20. A. Bax and S.S. Pochapsky, *J. Magn. Reson.*, 99 (1992) 638.
21. J. Jeener, B.H. Meier, P. Bachmann and R.R. Ernst, *J. Chem. Phys.*, 71 (1979) 4546.
22. G. Morris and R. Freeman, *J. Am. Chem. Soc.*, 101 (1979) 760.
23. R.M. Lynden-Bell, J.M. Bulsing and D.M. Doddrell, *J. Magn. Reson.*, 55 (1983) 128.
24. M.S. Friedrichs, W.J. Metzler and L. Mueller, *J. Magn. Reson.*, 95 (1991) 178.
25. A. Bax, A.F. Mehlkopf and J. Smidt, *J. Magn. Reson.*, 35 (1979) 167.
26. M. Rance, O.W. Sørensen, G. Bodenhausen, G. Wagner, R.R. Ernst and K. Wüthrich, *Biochem. Biophys. Res. Commun.*, 117 (1983) 479.
27. O.W. Sørensen, M.H. Levitt and R.R. Ernst, *J. Magn. Reson.*, 55 (1983) 104.
28. M.H. Levitt and R.R. Ernst, *J. Chem. Phys.*, 83 (1985) 3306.
29. C. Griesinger, O.W. Sørensen and R.R. Ernst, *J. Am. Chem. Soc.*, 107 (1985) 6394.
30. L. Mueller, *J. Magn. Reson.*, 72 (1987) 191.
31. L. Braunschweiler and R.R. Ernst, *J. Magn. Reson.*, 53 (1983) 521.
32. A. Bax, D.G. Davies and S.K. Sarkar, *J. Magn. Reson.*, 63 (1985) 230.
33. A.J. Shaka, C.J. Lee and A. Pines, *J. Magn. Reson.*, 77 (1988) 274.

34. M. Kadkhodaie, O. Rivas, M. Tan, A. Mohebbi and A.J. Shaka, *J. Magn. Reson.*, 91 (1991) 437.
35. R. Bazzo and J. Boyd, *J. Magn. Reson.*, 75 (1987) 452.
36. S. Macura, W.M. Westler and J.L. Markley, *Methods in Enzymology*, 239 (1994) 106.
37. K. Wüthrich, *NMR of Proteins and Nucleic Acids*. John Wiley and Sons, New York, 1986.
38. D. Neuhaus and M. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*. VCH, New York, 1989 .
39. L. Mueller, *J. Am. Chem. Soc.*, 101 (1979) 4481.
40. A. Bax, R.H. Griffey and B.L. Hawkins, *J. Am. Chem. Soc.*, 105 (1983) 7188.
41. M.R. Bendall, D.T. Pegg and D.M. Doddrell, *J. Magn. Reson.*, 52 (1983) 81.
42. G. Bodenhausen and D.J. Ruben, *Chem. Phys. Lett.*, 69 (1980) 185.
43. A.G. Redfield, *Chem. Phys. Lett.*, 96 (1983) 537.
44. E.R.P. Zuiderweg, *J. Magn. Reson.*, 86 (1990) 346.
45. A. Bax, R.H. Griffey and B.L. Hawkins, *J. Magn. Reson.*, 55 (1983) 301.
46. A.S. Edison, F. Abildgaard, W.M. Westler, E.S. Mooberry, and J.L. Markley, *Methods in Enzymology*, 239 (1994) 3.

## Chapter 2

# Multidimensional NMR of Macromolecules

LUCIANO MUELLER and N. VASANT KUMAR

### 1. INTRODUCTION

This chapter presents an outline of the current state of multidimensional nuclear magnetic resonance spectroscopy and its impact on the characterization of biological macromolecules. Recent advances, particularly in heteronuclear, multidimensional NMR, made it possible to elucidate three-dimensional structures of increasingly complex proteins with remarkable precision. These developments dramatically elevated the importance of NMR in biomedical research, where many targeted proteins or nucleic acids, or protein complexes were not amenable to conventional two-dimensional proton NMR methods. Multidimensional NMR spectroscopy has evolved into a powerful, versatile analytical tool for the characterization of biomolecular structures in solution. After the conception of the basic principle of 2D NMR spectroscopy in 1972 by J. Jeener [1], Richard R. Ernst's visionary and pioneering work [2–5] laid the foundation for the geometric growth of a vast array of two [6–24] and higher dimensional NMR experiments [25–44] which are applied to solve an ever growing array of analytical and structural problems. The impressive growth of NMR as a research tool has been facilitated by dramatic and concurrent advances in computer technology, which provided the tools for multidimensional data processing, impressive advances in NMR spectrometer design and revolutionary breakthrough in biotechnology [45–48] which has facilitated the preparation of samples of wild type and mutant macromolecules and the incorporation of stable isotope labels into biomolecules.

The development of homonuclear techniques which establish through-bond [3] and through-space [49] interactions among protons, discussed in detail in Chapter 1, made it possible to elucidate structures

of globular proteins by 2D NMR [50]. In 1983 Wüthrich and co-workers demonstrated that three-dimensional conformations of proteins can be derived from NMR [51,52] using the distance geometry approach [53, 54], and that the solution conformations of proteins indeed agreed to a large extent with structures obtained by X-ray crystallography. These results not only elevated NMR to an important tool for structure elucidation but it also validated and in some instances facilitated the X-ray crystallographic analysis where medium resolution NMR derived structures serve as a starting point of X-ray based structure refinements. To date, NMR has become almost universally accepted as a tool for elucidating structures of organic molecules.

The ability to incorporate magnetically active  $^{13}\text{C}$  and  $^{15}\text{N}$  spin-1/2 nuclei into biomolecules (proteins, nucleic acids and carbohydrates) by modern biotechnology methods [55,56] has stimulated the development of numerous heteronuclear three-dimensional and four-dimensional NMR experiments in the few years [39]. In a typical multidimensional NMR experiment, nuclear spin magnetization is shuffled among protons,  $^{15}\text{N}$  and  $^{13}\text{C}$  nuclei. In these spectra NMR resonances are distributed over an  $n$ -dimensional cube spanned by three to four spectral axes, one of which is almost always proton (acquisition domain) and the indirectly detected domains can be any combination of proton, or heteronuclei (e.g.  $^{13}\text{C}$ ,  $^{15}\text{N}$ ) frequencies. This spread of peaks over three or more dimensions serves both to increase spectral resolution and to reveal detailed connectivity paths among protons, nitrogens and carbons. These intrinsic features of multidimensional spectra make it possible to study considerably higher molecular weight proteins [57,58]. To date, NMR resonance assignments have been reported on proteins that are in excess of 30 kDa [59–62] and solution structures have been determined on proteins that are larger than 20 kDa [63–65].

Thanks to these dramatic developments, multidimensional NMR has become an indispensable tool in biomedical research. It may become possible in the near future to solve structures of larger biomolecules in the molecular weight range of perhaps up to 50 kDa and beyond using heteronuclear multidimensional NMR methods. Clever use of selective  $^{15}\text{N}$  or  $^{13}\text{C}$  labels [55,66] or the dilution of protons by deuterons increases the molecular weight limit even further, as shown towards the end of this chapter. Hence, NMR is a uniquely qualified method to facilitate structure-based drug design in biomedical research by providing detailed structural information on complexes between prospective inhibitory molecules and enzymes or soluble domains of receptors.



## 2. PRINCIPLES OF MULTIDIMENSIONAL NMR SPECTROSCOPY

This section is kept relatively brief since the principles of multidimensional NMR have been reviewed extensively elsewhere and the two-dimensional approach has already been discussed in the previous chapter. Multidimensional experiments are straightforward extensions of 2D experiments whose building blocks are well known — they consist of an excitation period, an evolution period  $t_1$ , a mixing period  $\tau_m$ , and a detection period  $t_2$  [4]. This schematic representation of a multidimensional time-resolved experiment is depicted in Fig. 1. Any multidimensional experiment begins with a preparation period, where a desired non-equilibrium spin state is prepared. This non-equilibrium state subsequently evolves in the evolutionary interval  $t_1$ . The  $t_1$  interval is typically terminated by an rf pulse which marks the beginning of the mixing period where coherence/magnetization is exchanged among magnetic nuclei according to the selected mode. In a two-dimensional experiment the mixing period is followed by the detection period, where transverse magnetization (typically protons) is detected.

This scheme can be easily extended to arbitrary dimensionality by the concatenation of a string of evolutionary periods  $t_j$  (where  $j = 1, 2, 3$  etc.), each followed by a mixing period as shown in Fig. 1b. During a mixing period nuclear spins are induced to exchange magnetization either by through-bond coherence transfers via scalar couplings, through space transfers via dipolar relaxation or chemical exchange depending on the type of chosen stirring of detected magnetic nuclei. Figure 1c depicts an example of a 4D extension of a two-dimensional NOE experiment, namely, a  $^{13}\text{C}$ – $^{15}\text{N}$  correlated 4D NOESY. The benefits of  $^{15}\text{N}$  or  $^{13}\text{C}$  correlated multidimensional experiments such as the depicted 4D CN-NOESY are two fold [67]. Cross peaks in the corresponding multidimensional experiment are much more uniquely defined since the resonances of both the donor and acceptor protons are dispersed along a second heteronuclear frequency domain. Hence, unique assignments of NOEs result for all protons that give resolved cross peaks in the corresponding two-dimensional  $^1\text{H}$ – $^{13}\text{C}$  or  $^1\text{H}$ – $^{15}\text{N}$  spectra [8,11,12,68]. Second, the relatively large and almost conformation independent magnitude of one bond coupling constants between  $^1\text{H}$ – $^{13}\text{C}$ ,  $^1\text{H}$ – $^{15}\text{N}$  and  $^{13}\text{C}$ – $^{13}\text{C}$  pairs greatly facilitates rapid transfer of coherences among the hetero atoms, ensuring relatively high conformation independent sensitivity in these experiments.

Homonuclear multidimensional experiments do not require isotopically enriched samples. While heteronuclear multidimensional spectra

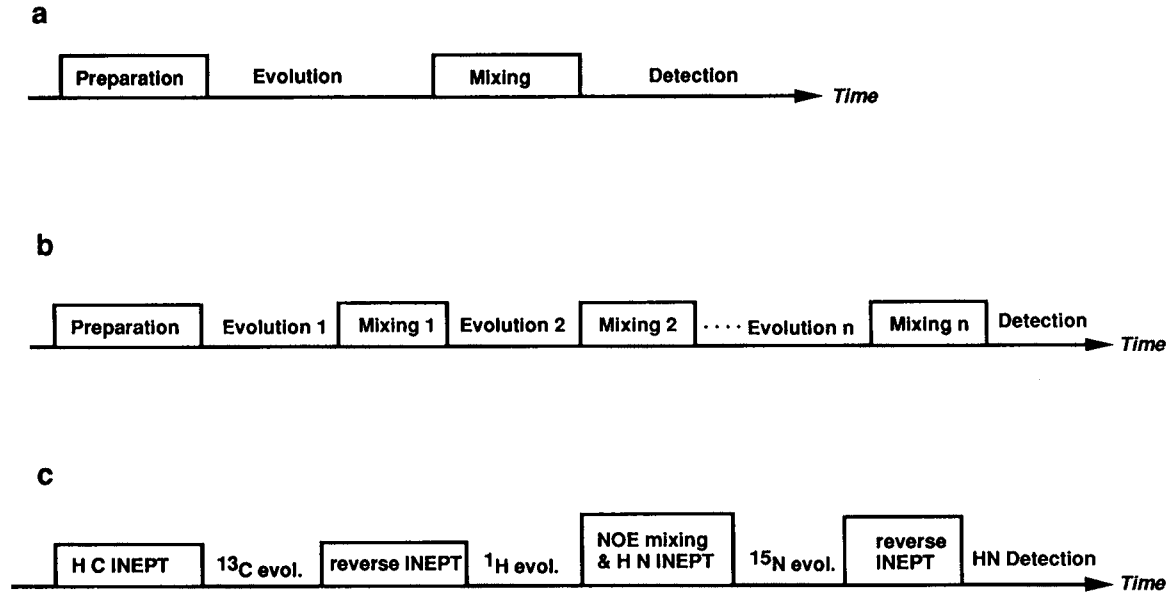


Fig. 1. (a) General building blocks of a two-dimensional NMR experiment; (b) building blocks of an n-dimensional experiment; (c) building blocks of a four-dimensional  $^{13}\text{C}$ - $^{15}\text{N}$  correlated NOESY experiment.

contain about the same number of cross peaks as their two-dimensional counterparts, homonuclear nD spectra are considerably more complex than their corresponding 2D spectra [29,69,70]. This is due to the fact that in heteronuclear experiments it is possible to transfer magnetization from a proton to a single hetero-atom (e.g.  $\text{HN} \rightarrow {}^{15}\text{N}$ ), whereas in a homonuclear mixing period, magnetization from every proton is splattered among multiple partners. Therefore, homonuclear multidimensional spectra provide in principle more information but by the same token suffer from considerable complexity. In addition, homonuclear nD experiments tend to suffer from low sensitivity, because magnetization exchange must rely in part on relatively weak and strongly conformation dependent proton-proton J-couplings [71–73]. For these reasons, homonuclear multidimensional NMR has thus far been restricted to solving structural problems in relatively low molecular weight proteins where the proton line widths are comparable to homonuclear scalar couplings [74–83].

In light of this, it is understandable that the bulk of applications of the multidimensional NMR approach has occurred with heteronuclear methods. For this reason, we restrict the following discussion to heteronuclear methods. These techniques require labeling of compounds with stable isotopes. Fortunately, incorporation of stable isotope labels in protein can be achieved by a growing array of molecular biology methods as described in Chapter 3 and in Refs. [45–47].  ${}^{13}\text{C}$  and  ${}^{15}\text{N}$ -labeling is most feasible if properly folded proteins can be derived from a bacterial expression system [55,56]. Recently, successful incorporation of stable isotope labels in proteins was reported in non-bacterial expression [84–89]. Novel procedures have also been developed for the incorporation of  ${}^{13}\text{C}$  and  ${}^{15}\text{N}$  isotopes in nucleic acids [90–92]. Finally, an exciting novel technology is emerging which permits efficient type specific isotope labeling of protein by cell-free synthesis [93].

Most heteronuclear multidimensional NMR experiments can be viewed as straightforward extensions of two-dimensional experiments where resonances of donor or acceptor protons or both are dispersed along additional frequency domains. For example, in a 3D spectrum all peaks which were previously compressed in a 2D plane are now dispersed over a 3D cube, giving rise to a considerably enhanced resolution as illustrated in Fig. 2.

Note that typically the number of cross peaks in a heteronuclear 3D or 4D spectrum is identical to the number of cross peaks found in a corresponding 2D spectrum provided that the peaks are all resolved.

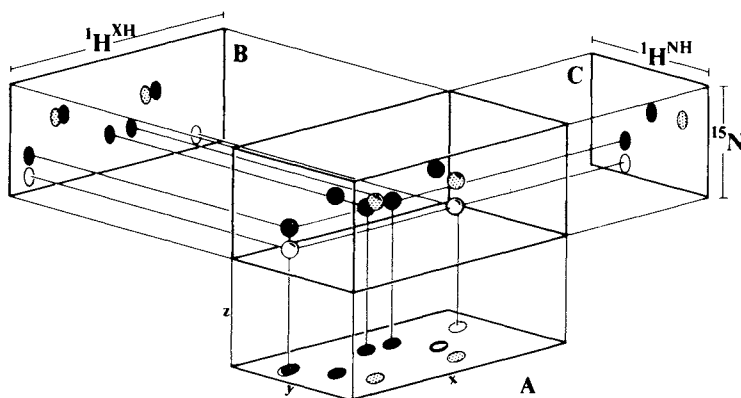


Fig. 2. An illustration of a  $^{15}\text{N}$ -correlated 3D NMR spectrum (center) with its three orthogonal projections. Projection A corresponds to a homonuclear 2D spectrum, whereas projections B and C correspond to  $^{15}\text{N}$ -correlated 2D and HN-HSQC or HMQC spectrum respectively [94].

Figure 3 illustrates the improved resolution in 2D NOESY spectrum provided by a 3D extension. Figure 3a depicts three 2D cross-section of a 3D  $^{15}\text{N}$ -resolved NOESY in apokardarcidin (a 114 residue globular protein) [88]. These cross-sections were taken at the  $^{15}\text{N}$ -resonances of three sequential amino acid residues i.e. Gly62, Thr63 and Thr64. Sequential NOEs from the amide proton of Thr64 to  $\text{H}_\alpha$ ,  $\text{H}_\beta$  and  $\text{H}_\gamma$ -protons of Thr63 are clearly resolved. In addition, the amide proton of Thr63 reveals additional well-resolved connectivities to the two  $\text{H}_\alpha$ -protons of Gly62. These connectivities would have been much more difficult to establish in a corresponding 2D NOESY spectrum as evidenced by the crowded map depicted in Fig. 3b, which shows the superposition of all  $^{15}\text{N}$ -plane of the 3D  $^{15}\text{N}$ -correlated NOESY spectrum. Furthermore, the amide protons of Gly62 and Thr63 are almost degenerate (8.67 ppm versus 8.60 ppm), whereas the corresponding amide  $^{15}\text{N}$ -resonances are well-resolved.

### 3. DELINEATION OF INTER-NUCLEAR CONNECTIVITY NETWORKS

#### 3.1. Introduction

Before any structural information can be extracted from NMR data, the vast number of magnetic resonance peaks one encounters in a

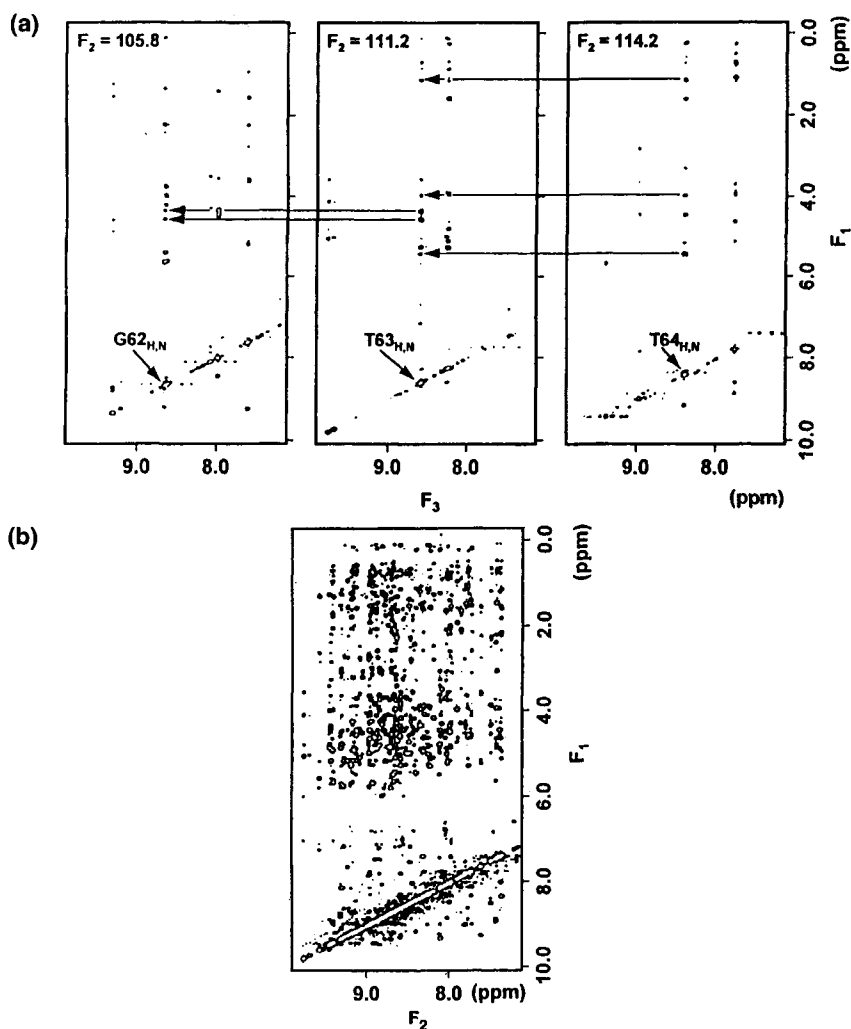


Fig. 3. 3D  $^{15}\text{N}$ -correlated NOESY [28,95] spectrum of apokedarcidin. Sample preparation as described in [88], sample concentration 2 mM, temperature = 20°C. Spectral widths and evolution times were as follows: SW1( $^1\text{H}$ ) = 7000 Hz,  $t_{1\text{max}}$  = 27.4 ms; SW2( $^{15}\text{N}$ ) = 1600 Hz,  $t_{2\text{max}}$  = 10.6 ms, SW3(NH) = 7000 Hz,  $t_{3\text{max}}$  = 70 ms. Number of scans per  $(t_1, t_2)$ -value = 8,  $t_2$ -evolution was performed in constant time, and pulsed field gradients (pfgs) were used to suppress undesired coherence paths (unpublished sequence). The duration of the experiment was 35 hours. The  $t_2$ -interferogram was doubled by mirror image complex linear prediction using the algorithm by Zhu and Bax [96]. (a)  $(F_1, F_3)$ -cross sections at the  $^{15}\text{N}$  resonance maxima of three sequential amino acid residues. (b) Skyline projection of the entire 3D spectrum along the  $F_2$  ( $^{15}\text{N}$ ) axis.

biological macromolecule must first be assigned to individual atomic nuclei. In the pre-Fourier spectroscopy era magnetic resonance assignments involved selective irradiation of individual resonances in a 1D spectrum [97,98]. This was time-consuming and selectivity was difficult to come by in a crowded spectrum. In a multidimensional time domain experiment all resonances of interest are simultaneously excited, and considerable resolution is achieved by spreading resonance connectivities across a multidimensional cube (Fig. 3). Connectivities between groups of magnetic nuclei are achieved by propelling magnetic nuclei to exchange magnetization during a corresponding pulse sequence in a multidimensional NMR experiment. There are basically three mechanisms by which magnetic nuclei can exchange magnetization. First, magnetization can travel along chemical bonds via scalar or  $J$  spin-spin couplings [1,3]. Second, nuclei can exchange magnetization through space by means of cross-relaxation which is induced by fluctuating dipolar fields [99,100]. Third, magnetic nuclei can swap places and thereby carry magnetization through space [49] (chemical exchange); for example, a polarized proton from a nearby water molecule can trade places with a protein bound amide proton [101,102].

Modern sequential resonance assignment strategies make extensive use of heteronuclear multidimensional experiments that invoke through-bond magnetization transfers [4,33,50,103,104] (see Fig. 4). A discussion of the most prominent experiments of this class is presented in the next section. After completion of the sequential resonance assignments, i.e. once the resonance frequencies of most magnetic nuclei are known, one can proceed with the derivation of structural information from NMR data [39,50,105]. Alternatively, one can try to skip these through-bond correlation experiments and to pursue simultaneous extraction of resonance assignments and structural data directly from NOESY spectra, as proposed in a recent paper by Kraulis [106]. This novel approach appears very intriguing but it remains to be seen whether it proves to be more economical than the subsequently outlined approach. It is tempting to establish through-bond connectivities by a few well chosen through-bond experiments CBCANH [107], HA(CO)CANH [108], HNCAHA, HN(CO)CAHA [109–111], CBCA(CO)NH [38], HNCACB [112], HN(CA)HA [113,114] or the glycine selective HN(CA)HA [115], once one has gone through the trouble of preparing a [ $^{15}\text{N}$ ,  $^{13}\text{C}$ ]-labeled sample. Recently gradient enhanced versions of some of these experiments have been published i.e. the GE-HNCAHA [116] and the GE-HNCACB (3D and 4D) and the GE-HN(CO)CACB [59]. In addition,

Grzesiek et al. published an elegant 4D HN(COCA)HN-experiment [117], which establishes direct through-bond correlations between pairs of sequential amide groups. However, this last experiment requires triply [ $^{13}\text{C}$ ,  $^2\text{H}$ ,  $^{15}\text{N}$ ]-labeled protein samples.

### 3.2. Through-bond correlations

Through-bond connectivities between chemically remote pairs of magnetic nuclei can be established by a stepwise relay of magnetization along a chain of chemically linked magnetic nuclei by utilizing relatively large one-bond scalar coupling constants among hetero atoms. Most of these couplings are favorably large, i.e. exceeding the intrinsic resonance line widths in small to mid-size globular proteins ( $\leq 30$  kDa) (see Table 1).

There are basically two ways of shuffling magnetization among chemically bonded heteronuclei, namely by isotropic mixing [19,130–132] (TOCSY/Hartmann–Hahn) either homonuclear and/or heteronuclear or by INEPT-type transfers [11,15,133,134]. During isotropic mixing, strong coupling is induced among nuclei by an appropriate pulse-train which quenches differences in chemical shifts while leaving the spin–spin couplings of interest intact. In the INEPT transfer non-equilibrium states of the second kind of the form  $I_y S_z \sin(\pi J \tau)$  are created during the INEPT interval ( $90^\circ_y - 1/4J_{IS} - 180^\circ_x - 1/4J_{IS} -$ ) [133]. A  $90^\circ_x$ -pulse immediately following the INEPT interval converts the non-equilibrium two-spin state into  $I_z S_y \sin(\pi J \tau)$ , thereby relaying coherence from spin I to spin S. The general aim is to establish a web of through-bond connectivities among as many magnetic nuclei as possible by performing an appropriate set of multidimensional experiments. The most prominent experiments of this class are depicted in Figs. 4 and 5. The following paragraphs illustrate how this can be accomplished in peptides and proteins.

TABLE 1

One-bond coupling constants in Hz [73,127–129]

$^1J_{\text{CH}}$	128–150 ( $\text{sp}^3$ )	160–170 aromatics ( $\text{sp}^2$ )	
$^1J_{\text{NH}}$	89–94		
$^1J_{\text{CC}}$	34–40 (aliphatic)	54–60 ( $\text{C}_\alpha\text{CO}$ )	66 (aromatics)
$^1J_{\text{CN}}$	14.8–15.6 (N CO)	9–11 ( $\text{NC}_\alpha$ )	
$^2J_{\text{CN}}$	6.4–8.3		

### 3.2.1. Backbone connectivities involving amide protons

The basic foundation for the structure elucidation of globular proteins and nucleic acids from two and higher dimensional NMR experiments has been laid out to a considerable extent by Wüthrich and collaborators including Richard Ernst [50,51,135,136]. Although many important contributions have been made by others [54,137–145], Wüthrich and collaborators were the first to vigorously pursue and promote the potential of multidimensional NMR in structural biology [50]. In order to overcome spectral complexities in biological macromolecules, Wüthrich and coworkers developed novel strategies for spectral resonance assignments in these systems. They showed that it is uneconomical to attempt complete assignments in one giant leap. Instead, it is more efficient to divide the resonance assignments of proteins into several substeps. Preferably, resonances are first grouped into generic amino acid residues or through-bond coupled spin systems. Next, these generic residues are assigned to a particular amino acid residue types. Finally, these generic type specifically assigned residues are linked to one unique location in amino acid sequence of a given protein. Further refinement of the resonance assignment strategy is necessary in higher molecular weight proteins. There, it is often preferable to proceed with the sequential resonance assignment once the NH protons are uniquely linked to their own  $H_\alpha$  and  $H_\beta$  protons (NABs) [146]. Once a considerable portion of the backbone assignments have been tentatively established, it is recommended that the assignment be extended further down the amino acid sidechains.

The first step in the spectral assignments of [ $^{15}\text{N}$ ,  $^{13}\text{C}$ ]-labeled proteins is to sequentially link as many NH-groups as possible utilizing heteronuclear triple resonance experiments which establish through-bond connectivities across peptide bonds [57]. The most commonly used experiments of this kind, outlined in Fig. 4, are: CBCANH [107], HA(CO)CANH [108], HNCAHA, HN(CO)CAHA [109–111], CBCA(CO)NH [38], HNCACB [112], HN(CA)HA [113,114] or the glycine selective HN(CA)HA [115]. All of these experiments rely on heteronuclear one-bond couplings which are only moderately dependent on conformation. For this reason, these experiments give rise to relatively uniform cross-peak intensities. This uniformity in peak heights tends to facilitate the spectral analysis.

All thirteen experiments depicted in Fig. 4 link amide HN,  $^{15}\text{N}$  proton–nitrogen pairs to chemically adjacent magnetic nuclei. First and foremost, it is important to note that it is usually not necessary to perform all those thirteen multidimensional experiments in order to



achieve complete sequential resonance assignments along the peptide backbone. In fact, in most small to midsize proteins the first five experiments in Fig. 4 provide sufficient information to sequentially link

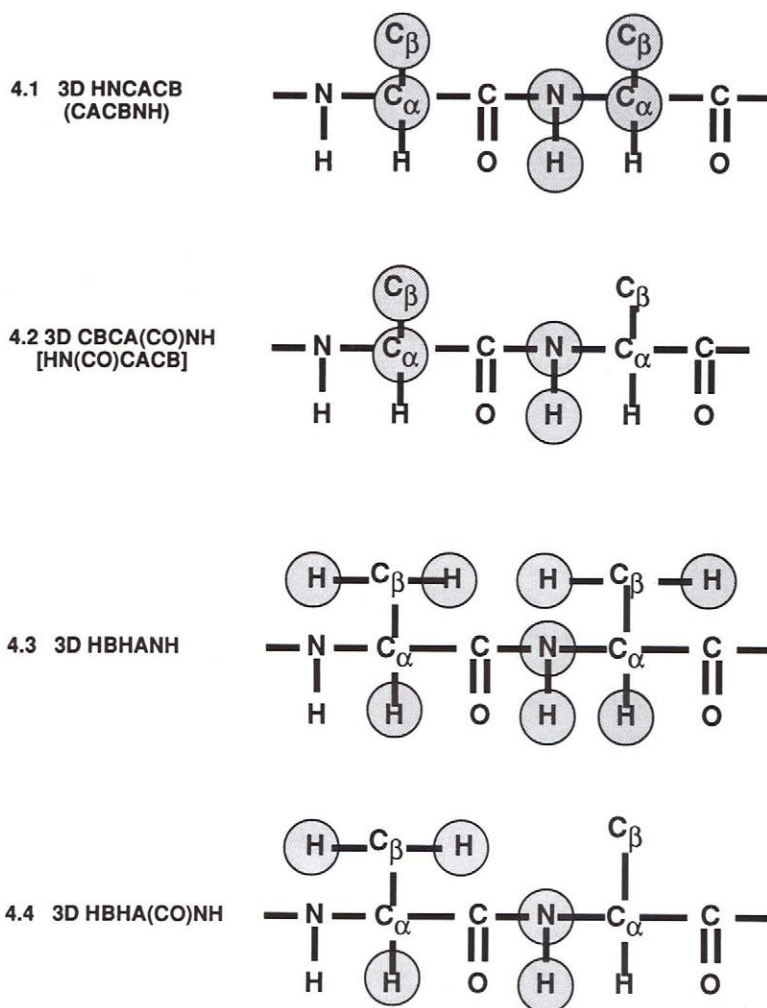


Fig. 4. Left column: Names of major triple resonance experiments which are used to perform sequential resonance assignments if [ $^{15}\text{N}$ ,  $^{13}\text{C}$ ]-labeled samples are available. Right column: diagram of two sequential amino acid residues, connected nuclei are highlighted by shaded circles.

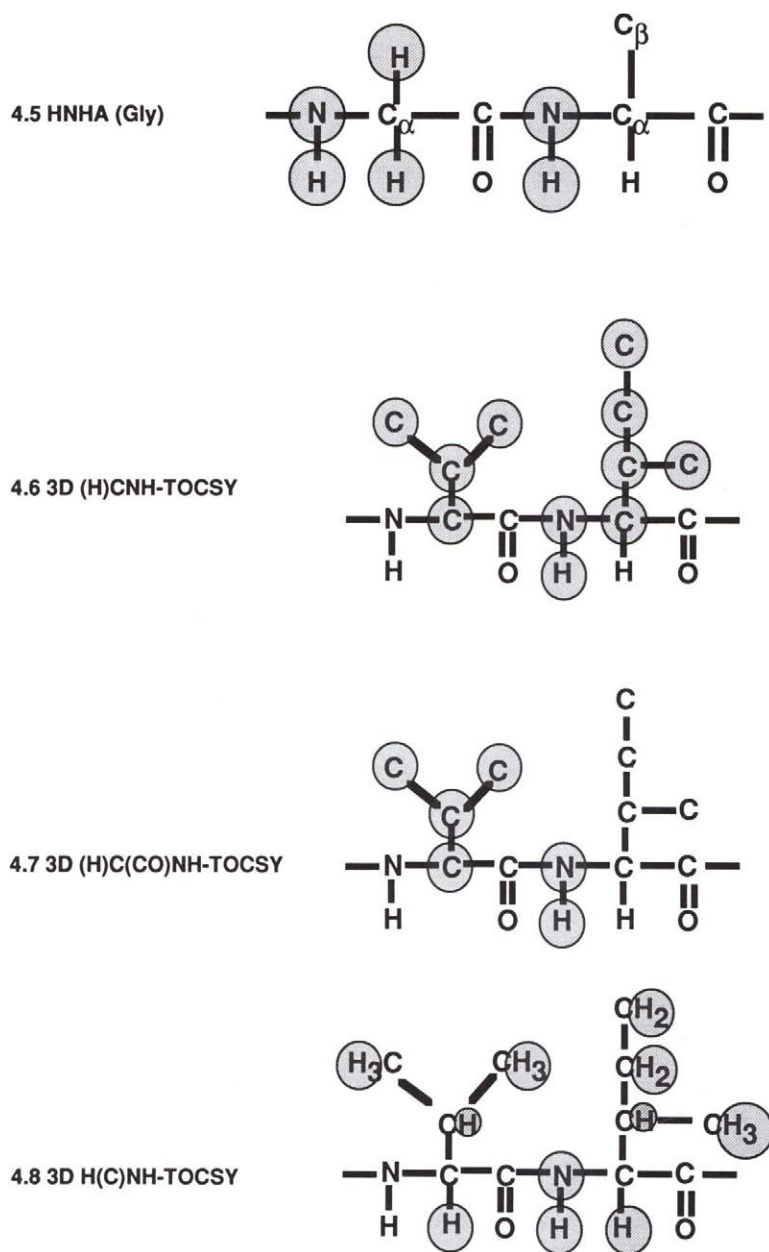


Fig. 4 (continued).

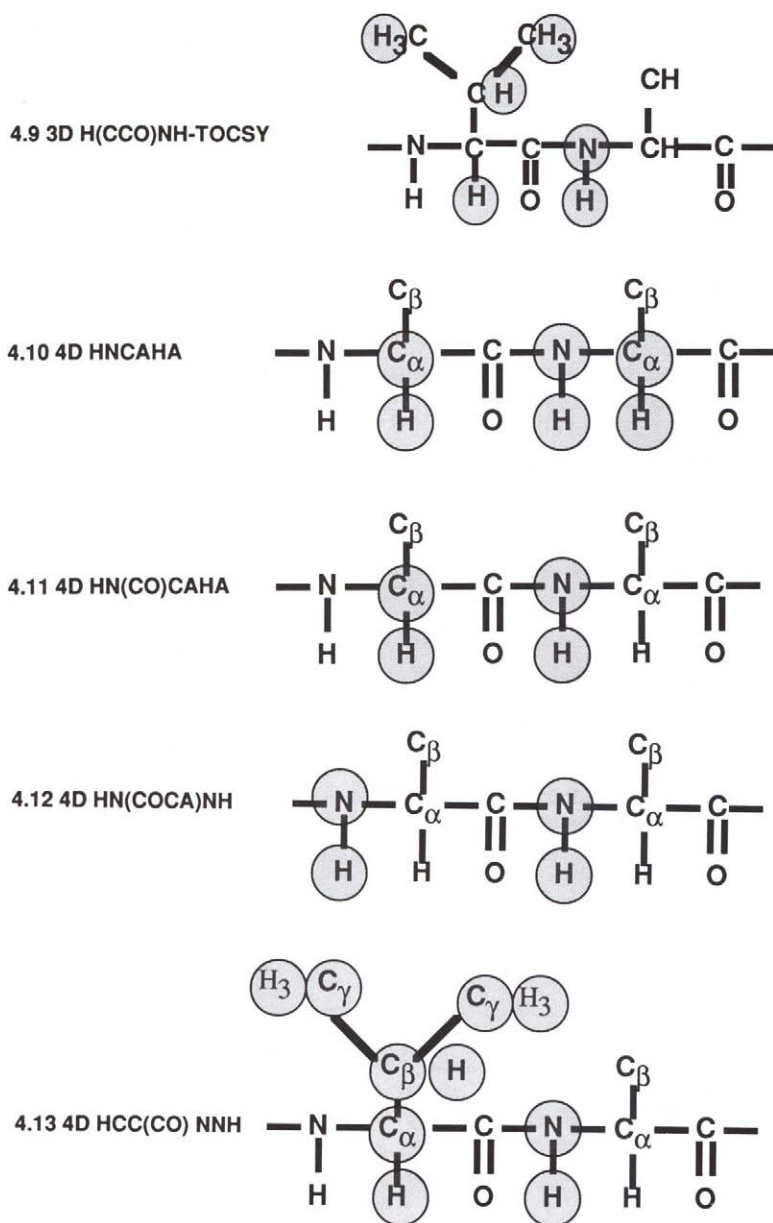


Fig. 4 (continued).

all the backbone NH-resonances including the  $C_\beta$  and  $H_\beta$  resonances of the entire protein excluding oligo proline stretches. The first experiment, i.e. the HNCACB (CBCANH) [42,107], links the NH's to intra residual, and in favorable cases sequential ( $i-1$ ),  $C_\alpha$  and  $C_\beta$  carbons. Figure 5 depicts the detailed pulse sequence of HNCACB experiment. The CO-edited analog of the HNCACB experiment, i.e. the HN(CO)CACB [59] or CBCA(CO)NH [38], correlates the NHs to sequential  $C_\alpha$  and  $C_\beta$ -carbons exclusively by selecting coherence paths which relay magnetization via carbonyl carbons across peptide bonds.

---

*Opposite:* Fig. 5. Pulse sequence of HNCACB-experiment [42]. Thin vertical lines represent  $90^\circ$  pulses while the thicker ones represent  $180^\circ$  pulses, hatched rectangles represent composite  $180^\circ$  pulses, hatched half ellipsoids represent frequency shifted g3-pulses [147], the solid half ellipsoid at  $t_1/2$  is a frequency shifted Gaussian pulse. Optional deuterium decoupling during the  $\delta_2$  and the  $t_1$  evolution periods is achieved with a WALTZ-16 [148] pulse train with an amplitude of about 1–2 kHz centered at 3.5 ppm. Pulse intervals are as follows:  $\tau' = 2.5$  ms,  $\tau = 2.71$  ms,  $\Delta_2 =$  sum of durations of selective water-flip-back pulse; gradient pulse and gradient recovery delay,  $\delta_1 = 11$  ms,  $\delta_2 = 6.9$  ms {HNCB} or  $\delta_2 = 3.45$  ms {HNCACB},  $\zeta = 0.7$  ms. Phase table: (a) HNCA:  $\phi_1 = \{8(y), 8(-y)\}$ ,  $\phi_2 = -x$  plus small angle phase adjustment for optimal flip-back of  $H_2O$  magnetization,  $\phi_3 = \{4(x), 4(-x)\}$ ,  $\phi_5 = \{y\}$ ,  $\phi_6 = \{2(y), 2(-y)\}$ ,  $\phi_{f1}, \phi_{f2}$  are set to x or y to collect the respective components of the  $t_1, t_2$ -interferograms,  $\phi_{rec} = \{x, 2(-x), x, -x, 2(x), 2(-x), 2(x), -x, x, 2(-x), x\}$ , (b) HNCACB or HNCB:  $\phi_1 = \{16(y), 16(-y)\}$ ,  $\phi_2 = -x$  plus small angle phase adjustment for optimal flip-back of  $H_2O$  magnetization,  $\phi_3 = \{8(x), 8(-x)\}$ ,  $\phi_5 = \{4(x), 4(-x)\}$ ,  $\phi_6 = \{2(y), 2(-y)\}$ ,  $\phi_{f1}$  is set to x or y to collect the respective component of the  $t_1$ -interferogram, the x-component of the  $t_2$ -interferogram is collected by setting  $\phi_{f2} = x$ , the y-component of the  $t_2$ -interferogram is obtained by setting  $\phi_{f2} = -x$  and by inverting the sign of the G5 amplitude,  $\phi_{rec} = \{2(x, -x), 4(-x, x), 2(x, -x), 2(-x, x), 4(x, -x), 2(-x, x)\}$ .  $^1H$ ,  $^{15}N$ , and  $^{13}C\alpha/\beta$  pulses are applied at 4.8, 121 and 45 ppm with field strengths of 25, 7.1 and 17.9 kHz respectively. The WALTZ-16 decoupling was utilized for both  $^1H$  and  $^{15}N$ -decoupling using RF field strengths of 5.9 and 1.3 kHz. Phase sensitive spectra were obtained by using STATES-TPPI-type acquisition [149]. The magnetic field gradient pulses are set to: G1 = (8 Gauss/cm, 0.4 ms), G2 = (16 G/cm, 1.0 ms), G3 = (14 G/cm, 0.8 ms), G4 = (6 G/cm, 1.0 ms), G4' = (12 G/cm, 1.0 ms), G5 = ( $\pm$  41.6 G/cm, 2.5 ms), G6 = ( $-$ 31.5 G/cm, 0.3 ms), G7 = (31.5 G/cm, 0.3 ms), G8 = (41.12 G/cm, 0.25 ms). In the HNCA experiment pulses and intervals within the dashed boxes are eliminated, SW1 is set to  $< 4500$  Hz, and the 1st  $t_1$ -value is set to  $1/SW1$  to accommodate the Gaussian CO-decoupling pulse which is set to 180 to 200 ms and the 1st point in the  $t_1$ -interferogram is reconstructed by linear prediction. In the HNCACB or HNCB collection mode, SW1 is set to 9000 Hz or a value that covers the entire aliphatic carbon region, the optional CO-decoupling pulse narrows  $C_\alpha$  (HNCACB) and  $C_\beta$ s of Asn, Asp, the 1st  $t_1$ -value is set to  $2/SW1$ , and the first two points in the  $t_1$ -interferogram need to be reconstructed by linear prediction, alternatively in the HNCB collection mode the CO-decoupling pulse is of lesser importance and may be omitted, thereby the first  $t_1$ -interval is preferably set to  $1/(2*SW1)$  (half-dwell) [150].

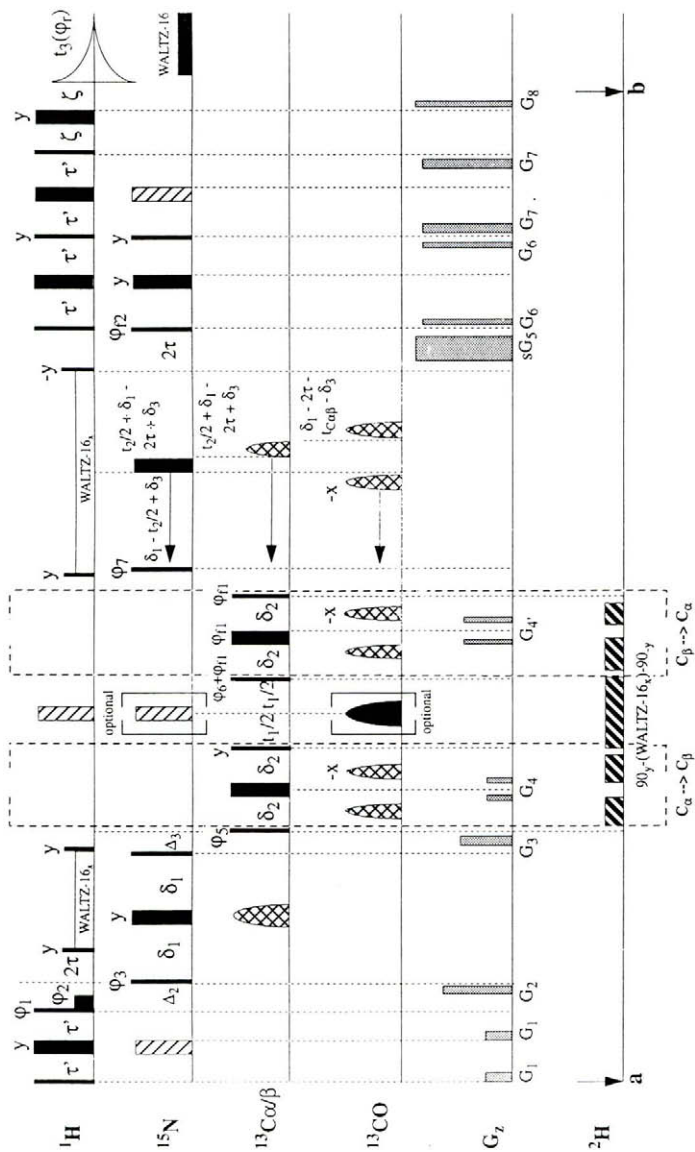


Fig. 5. Caption opposite.

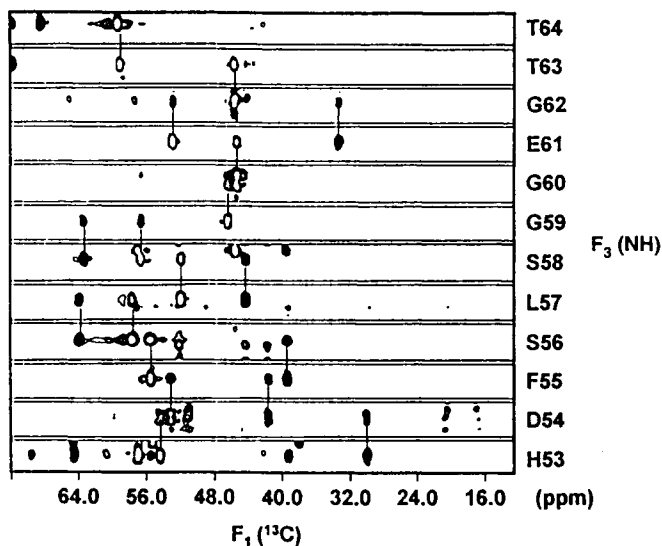


Fig. 6. (a) HNCACB-spectrum of apokedarcidin [88]. Cross peaks to  $C_{\alpha}$ -carbons have positive sign (open ellipsoids), and cross peaks to  $C_{\beta}$ -carbons have negative sign (solid ellipsoids). Sample conditions as described in Fig. 3. Spectral widths and evolution times were as follows:  $SW1(^1H) = 9000.9$  Hz,  $t_{1max} = 4.4$  ms;  $SW2(^{15}N) = 1600$  Hz,  $t_{2max} = 18.8$  ms,  $t_{3max} = 59$  ms. number of scans per  $(t_1, t_2)$ -value = 32. The  $t_1$ -interferograms were extended one-fold by complex linear prediction [96]. Duration of experiment 47 hours.

The editing effect of the carbonyl assisted experiment 4.2 (Fig. 4) permits clear distinction between intra-residual and sequential through-bond NH to  $^{13}C$  correlations. Examples of HNCACB and CBCA(CO)NH spectra are depicted in Figs. 6a and 6b respectively. Figure 6a depicts a series of cross-sections, which were taken at the  $^{15}N$ -resonance frequencies of a series of sequential residues in apokedarcidin. The depicted panels show cross peaks between HN-resonances of residue  $i$  to  $C_{\alpha}$  and  $C_{\beta}$  resonances of residues  $i$  and  $i-1$ . Note that the sequential cross peaks are of weaker intensities and cross peaks to  $C_{\alpha}$  and  $C_{\beta}$ -carbons are of opposite sign — a feature that facilitates identification of  $C_{\beta}$ -resonances particularly in serines and threonines. Cross-section through a corresponding CBCA(CO)NH spectrum of apokedarcidin neatly confirms all sequential linkages between NH( $i$ ) and  $C_{\alpha}(i-1)$  and  $C_{\beta}(i-1)$  as seen in Fig. 6b.

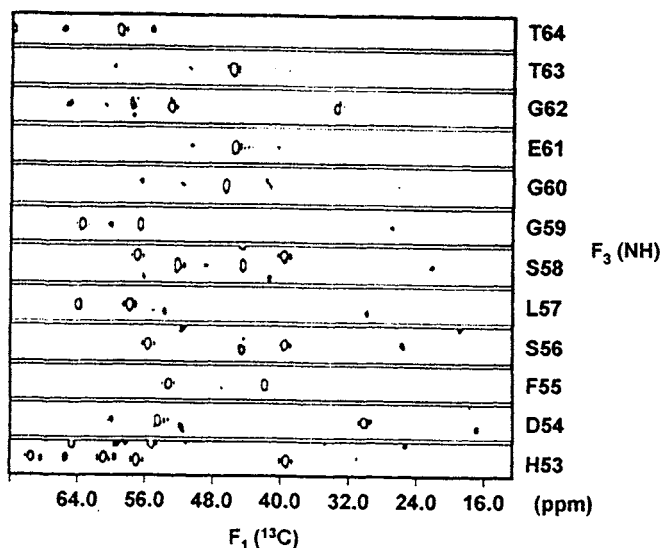


Fig. 6. (b) CBCA(CO)NH [38]-spectrum of apokedarcidin. Sample conditions as described in Fig. 3. Spectral widths and evolution times were as follows: SW1( $^1\text{H}$ ) = 9000.9 Hz,  $t_{1\text{max}}$  = 4.7 ms; SW2( $^{15}\text{N}$ ) = 1600 Hz,  $t_{2\text{max}}$  = 18.8 ms,  $t_{3\text{max}}$  = 59 ms. nt per ( $t_1, t_2$ )-value = 32. The  $t_1$ -interferograms were extended one-fold by complex linear prediction. Duration of experiment 47 hours. Depicted are a series of (F1,F3)-slices at the  $^{15}\text{N}$ -resonance maxima of the amino acid residues, which are shown in the far right column.

The HBHANH and HBHA(CO)NH experiments [40] (see experiments 4.3 and 4.4 in Fig. 4) are twin brothers of the HNCACB and CBCA(CO)NH experiment. They correlate the amides with the intra-residual and sequential  $\text{H}_\alpha$  and  $\text{H}_\beta$  protons. Experiment 4.5 (Fig. 4) [118] correlates geminal proton pairs to NH groups. This includes glycine  $\text{H}_\alpha$ 's, arginine  $\text{H}_\delta$ , asparagine and glutamine  $\text{H}_\beta$  and  $\text{H}_\gamma$  respectively. Hence, experiments 1–5 in Fig. 4 establish all intra-residual and sequential NAB pattern along the peptide chain except for polyproline stretches. This permits almost complete sequential resonance assignments up to the  $\text{C}_\beta$   $\text{H}_\beta$ s, except in rare cases of extreme resonance overlaps or insufficient sensitivity and of course connectivities to prolines which lack amide protons. It might not have gone unnoticed that the dimensionality of all of these five experiments is less than four. Restriction of the dimensionality to three provides superior sensitivity, better definition of resonance frequencies (more spectral points per

ppm), and it permits the collection of a larger number of experiments in a given time-period relative to 4D analogs. However, some of these restrictions have been alleviated with the advent of pulsed field gradients for coherence path selection, which can reduce the minimum number of scans per  $t_1$  value to two [59,116]. Allocating a maximum of two days (gradient enhanced versions of these experiments can be recorded in less than 24 hours each) per experiment suggests that ten days (less than five days with the availability of pulsed field gradient accessory) of data collection in a single protein sample is minimally required to complete the backbone resonance assignments. This is a modest amount of time compared to the time required to develop a bacterial expression system of a given protein and the preparation of a doubly labeled [ $^{15}\text{N}$ ,  $^{13}\text{C}$ ]-protein sample.

The C-NH-TOCSY [37,119], its CO-correlated brother [37,120,121], the C(CO)-NH-TOCSY (experiments 4.6 and 4.7 in Fig. 4), the  $^1\text{H}$ -correlated analogs (experiments 4.8 and 4.9 in Fig. 4) and the 4D combination of experiments 4.7 and 4.9 (Fig. 4) [43,109,116,122,123]. Experiment 4.13 (Fig. 4) in principle links all aliphatic sidechain carbons and protons to their intra-residual and sequential amide proton- $^{15}\text{N}$  pairs. Thereby, these neat experiments offer the potential of quickly wrapping up the remaining sidechain assignments in aliphatic amino acid residues. Unfortunately, all of these techniques with the exception perhaps of the HNCAHA (experiment 4.10 in Fig. 4) feature lower sensitivity than experiments 1-5 and may yield only partial assignments of sidechain resonances, particularly in higher molecular weight proteins. However, the experiments which utilize CC-TOCSY transfers are particularly useful in providing assignments in partially unfolded or in deuterated proteins which tend to feature long transverse relaxation times  $T_2$  [124]. In experiments 4.6 and 4.7 (Fig. 4) for example, after labeling the sidechain carbons in  $t_1$  a portion of the sidechain carbon magnetization is transferred to amide nitrogens via a concatenation of CC-TOCSY and subsequent CN-INEPT. Amide nitrogen frequencies are encoded in  $t_2$  and amide protons are finally detected in  $t_3$  under inverse  $^{15}\text{N}$ -decoupling. Hence, in a C-NH-TOCSY spectrum sidechain carbons are linked to amide groups. Again, its CO-assisted brother (experiment 4.7 in Fig. 4) permits clear distinction between intra-residual and sequential connectivities. Experiments 4.8 and 4.9 (Fig. 4) are closely related to experiments 4.6 and 4.7 (Fig. 4) by providing through-bond connectivities between sidechain protons and amide groups. In principle, the  $^{15}\text{N}$ -correlated HH double resonance experiment [24]



gives the same type of information as the H(C)NH-TOCSY (experiment 4.8) but with lower sensitivity.

Experiments 4.10–4.13 (Fig. 4) represent four-dimensional through-bond correlation experiments. Hence, they are fairly time-consuming, particularly if one wishes to achieve a halfway decent resolution in all four dimensions, unless one trusts the resolving power of linear prediction or maximum entropy resolution enhancement techniques. In many instances 3D analogs provide the desired through-bond correlations in a shorter time period. Experiment 4.12 (Fig. 4) [117,125] has been designed to establish direct connectivities between sequential amide groups. This experiment works particularly well in deuterated protein samples where the bulk of the  $H_\alpha$ -protons are replaced by deuterons, because in this experiment the magnetization spends a considerable amount of time as transverse  $C_\alpha$ -carbon magnetization during the polarization transfer from one NH to the other. The appeal of experiment 4.12 (Fig. 4) is its ability to reveal direct sequential linkages of amide groups, thereby permitting a straightforward sequential walk along the peptide backbone.

A detailed description of the coherence pathways in each of these 13 experiments, as depicted in Fig. 4, is beyond the scope of this chapter. Instead, we wish to present the coherence pathway in terms of the product operator formalism [126] for one representative example, namely the HNCACB experiment in the following few paragraphs. An additional benefit of this experiment beside providing sequential connectivity is the strong dependence of  $C_\alpha$  and  $C_\beta$ -carbon shifts on amino acid residue type. This feature renders this experiment particularly valuable in locating strings of sequentially linked carbon resonances on the amino acid sequence. For this reason, and to present a typical example of a through-bond experiment we describe this pulse sequence in greater detail.

The major coherence pathway in the HNCACB [42] sequence is

$$HN^i \rightarrow N^i \rightarrow C_\alpha^i \rightarrow C_\alpha^i / C_\beta^i \{t_1\} \rightarrow C_\alpha^i \rightarrow N^i \{t_2\} \rightarrow HN^i \{t_3\}$$

yielding intra residue correlations. A parallel transfer mediated by two-bond  ${}^2J_{N-C\alpha-1}$  yields the inter residue correlations shown in the pathway

$$HN^i \rightarrow N^i \rightarrow C_\alpha^{i-1} \rightarrow C_\alpha^{i-1} / C_{\beta 1}^{i-1} \{t_1\} \rightarrow C_\alpha^{i-1} \rightarrow N^i \{t_2\} \rightarrow HN^i \{t_3\}.$$

This scheme is a straightforward extension of the HNCA experiment which was originally proposed by Kay et al. [161]. In fact, elimination of the relay step highlighted by the parenthesis truncates the experiment into the original HNCA sequence. In this experiment, the carbonyl carbons are effectively decoupled by a selective homonuclear decoupling sequence (SEDUCE) [151,152]. An alternate experiment which provides equivalent information is the CBCANH [107]. At point (a) in the pulse sequence of Fig. 5 the coherence of interest is  $\beta^* \text{NH}_z \text{N}_y$ , where  $\beta = N\gamma h$ . During the subsequent  $2\tau$  period bilinear  $\text{HN}_z \text{N}_y$  coherence is converted into pure nitrogen transverse magnetization after which coherent proton decoupling is turned on. During the  $^{15}\text{N}$ - $^{13}\text{C}$  INEPT interval  $2\delta_1$  bilinear  $^{15}\text{N}$ - $^{13}\text{C}$  coherences evolve due to one-bond and two-bond  $^{15}\text{N}$ - $^{13}\text{C}$  couplings

$$\{\text{HN}_z \text{N}_y \rightarrow \text{N}_y \text{C}_{\alpha z} \sin[2\pi J_{\text{NC}\alpha} \delta_1]\}.$$

The pair of  $90^\circ$  pulses applied to  $^{15}\text{N}$  and  $^{13}\text{C}$  at time point (b) converts this coherence into  $\text{N}_z \text{C}_{\alpha y}$ . During the following INEPT interval  $\delta_2$  a portion of the  $\text{N}_z \text{C}_{\alpha y}$  coherence is relayed to  $\text{C}_\beta$  mediated by the one-bond  $^1\text{J}_{\text{CC}}$ -coupling (see Table 1)

$$\{\text{N}_z \sin(2\pi J_{\text{NC}\alpha} \delta_1) [\text{C}_{\alpha y} \cos(\pi J_{\alpha\beta} 2 \delta_2) - 2 \text{C}_{\alpha x} \text{C}_{\beta z} \sin(\pi J_{\alpha\beta} 2 \delta_2)]\}.$$

The  $90^\circ$  carbon pulse at time point (c) ( $t_1 = 0$ ) creates coherences which can be expressed as

$$\text{N}_z \text{C}_{\alpha y} \cos[\pi J_{\alpha\beta} 2 \delta_2] + 2 \text{N}_z \text{C}_{\alpha z} \text{C}_{\beta x} \sin[\pi J_{\alpha\beta} 2 \delta_2].$$

Subsequent  $t_1$  evolution and reverse INEPT yields at point (d) coherences of the form

$$\{\text{N}_z \text{C}_{\alpha z} \sin(2\pi J_{\text{NC}\alpha} \delta_1) [\cos^2(\pi J_{\alpha\beta} 2 \delta_2) \cos(\sigma_\alpha t_1) - \cos(\sigma_\beta t_1) \sin^2(\pi J_{\alpha\beta} 2 \delta_2)]\}.$$

Typically,  $\delta_2$  is set to 6.5 ms [ $1/(4 J_{\alpha\beta})$ ] which renders cross peaks to  $\text{C}_\alpha$  and  $\text{C}_\beta$ -carbons to be of comparable intensity but with opposite sign. This effect is illustrated in Fig. 6a which depicts a number of selected panels of a 3D HNCACB spectrum. There, cross peaks from NH to  $\text{C}_\beta$ -carbons have negative signs. This sign inversion of cross peaks is in general desirable because it permits unambiguous identification of

$C_\beta$ -carbons. Yet complications may arise in serine residues where  $C_\alpha$  and  $C_\beta$ -cross peaks may accidentally overlap which would result in mutual extinction of both cross peaks. Typically, weaker inter residue connectivities to carbons  $C_{\alpha i-1}$  and  $C_{\beta i-1}$  are the cause of a parallel coherence pathway which is mediated by the two-bond  $^2J_{\text{NiC}\alpha i-1}$  couplings. Alternatively, two experiments can be collected back-to-back, one with  $\delta_2$  set to 13 ms  $[1/(2 J_{\alpha\beta})]$  followed by an HNCA. As a result one would obtain an HNCB and an HNCA spectrum each of which featuring fewer better resolved cross peaks particularly for serine residues where  $C_\alpha$  and  $C_\beta$ -resonances have similar chemical shifts [153].

These weaker connectivities are useful in making sequential assignments, but a more efficient way of obtaining this information is provided by the CBCA(CO)NH experiment, where the relatively large  $^1J_{\text{NCO}}$  coupling (15 Hz) [73,127] is employed to relay magnetization across peptide bonds (see Fig. 6b).

### 3.2.2. Sidechain resonance assignments

The HNCACB and HN(CO)CACB experiments, which contribute to the establishment of main chain sequential correlations, fortunately also provide extensions into the sidechains by yielding  $C_\beta$  and  $H_\beta$  resonance assignments. In the following paragraphs, additional experiments which extend through-bond connectivities further down the amino acid sidechains are described. The most frequently used experiments in this class are listed in Fig. 7. The first member in this series is the well-known HCCH-COSY (Fig. 7a) experiment [154], which is a heteronuclear extension of the good old proton COSY [1,3]. It correlates directly bonded  $^1\text{H}$ - $^{13}\text{C}$  pairs to vicinal protons. The experiment features high sensitivity since the coherence transfer proceeds via three one-bond transfer steps utilizing the large one bond  $^1J_{\text{CH}}$  and  $^1J_{\text{CC}}$ -couplings [73]. The HCCH-TOCSY [155,156] (Fig. 7b) is analogous to the proton TOCSY [132] (HOHAHA) [19] experiment. It establishes through-bond correlations among all aliphatic protons and carbons. Fesik and coworkers recently published a 4D extension of the HCCH TOCSY experiment [157] (Fig. 7c). This 4D HCCH TOCSY gives rise to spectra in which each aliphatic  $^{13}\text{C}$ - $^1\text{H}$  pair to all other aliphatic  $^{13}\text{C}$ - $^1\text{H}$  pair within a residue. For example, the HC-spectral plane of residue *i* originating at the coordinate of a  $C_\alpha H_\alpha$ -pair shows cross peaks to all other aliphatic CH-groups within residue *i*. This is convenient, although the experiment is time consuming. In the time it takes to record a 4D HCCH TOCSY one can collect both the 3D HCCH COSY and TOCSY

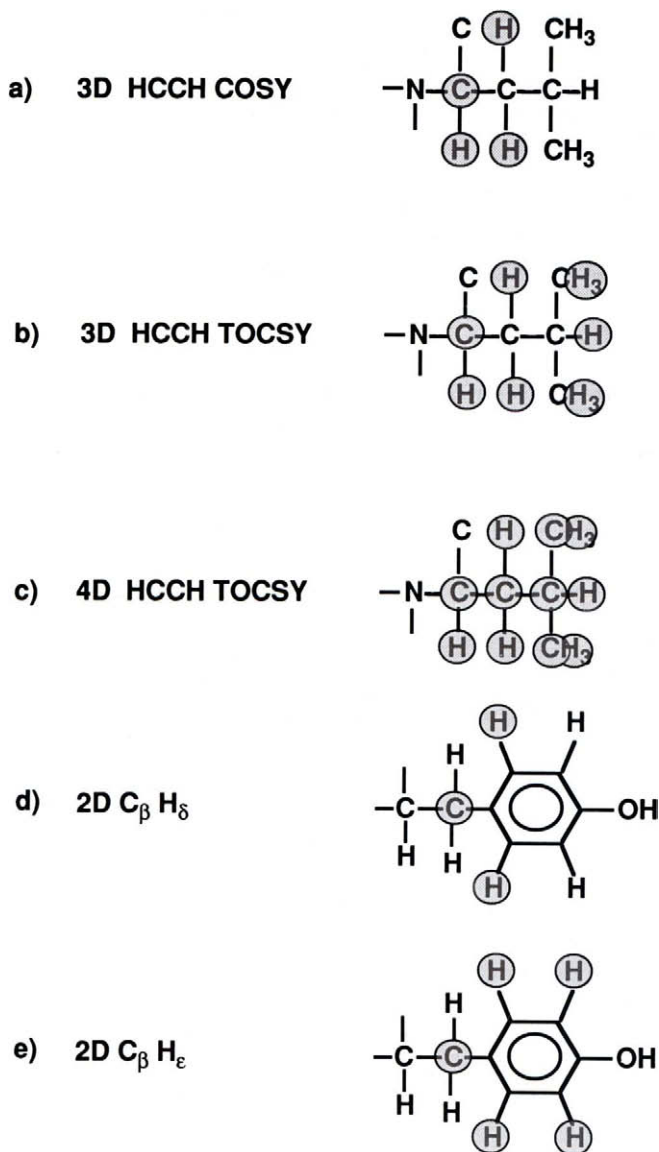


Fig. 7. Left column: names of major carbon-correlated double resonance experiments which are used to establish through-bond connectivities within amino acid side-chains. Right column: structures of the Leucine sidechain in a, b, c, and Tyrosine in d and e. Connected nuclei are highlighted by shaded circles.

experiments. These two experiments give additional information by revealing the distinction between a  $[H_\beta, H_\gamma]$  and an  $[H_\beta, H_\delta]$  unambiguously.

Aromatic sidechains pose a special challenge due to the large difference in resonance frequencies between  $C_\beta$ s and aromatic carbons. Typically,  $C_\beta$  carbons of aromatic residues resonate around 30–40 ppm, and aromatic  $C_\gamma$ ,  $C_\delta$ -carbons resonate around 130 ppm. This corresponds to a 13–15 kHz difference in high field spectrometers. A halfway decent  $C_\beta$  to  $C_\delta$  TOCSY transfer using a broadband spin locking sequence would require an rf field strength in excess of 13 kHz even with a very efficient spin locking sequence like FLOPSI-8 [158] for a duration of over 20 milliseconds. The savvy reader is well advised not to attempt this probe frying experiment.

Lewis Kay and coworkers achieved the desired through-bond connectivity by transferring  $C_\beta$ -magnetization to the aromatic sidechains by brute force relay and double relay steps in the 2D  $C_\beta$   $H_\delta$  and 2D  $C_\beta$   $H_\epsilon$  experiment [159] (Figs. 7d and 7e). These are the only 2D experiments that are discussed in this chapter. Three-dimensional extensions of these experiments are possible, but frequently the presented 2D analogs do the job of confirming aromatic sidechain assignments in a shorter time period. In these experiments  $C_\beta$ -carbon magnetization after evolving in  $t_1$  is transferred to aromatic carbons by a series of  $^{13}\text{C}$ – $^{13}\text{C}$  relay steps. Instead of letting the aromatic carbons evolve, which would be a bad idea, their coherence is immediately transferred to aromatic protons which are detected in  $t_2$ . Although these experiments have a relatively low sensitivity, they have, in our hands, revealed sidechain connectivities in all aromatic residues of human profilin which is a 139 residue protein. (Considerable improvement in sensitivity is obtained in these experiments by replacing the selective G3-pulse in the middle of the first  $\tau_c$ -delay ( $\Phi_6$ ) as depicted in Fig. 1 of Ref. [159] with a soft square pulse). Most recently Grzesiek and Bax proposed an even more efficient experiment for correlating  $C_\beta$ -carbon resonances with aromatic proton resonances employing a doubly rotating frame nutation isotropic mixing (AMNESIA) which establishes strong couplings between the  $C_\beta$  and the  $C_\gamma$ -carbons during the isotropic mixing and thereby enhancing the rate of magnetization transfer [160].

In addition to the experiments discussed thus far, a series of heteronuclear experiments have been developed which involve carbonyls more explicitly [161–168]. There, carbonyls evolve during a time domain and thereby give rise to spectra in which resonances are dispersed along the additional CO-frequency domain. This procedure enhances the resolution

since the carbonyls tend to be as well dispersed as the amide nitrogens and also feature long  $T_2$ -values. This procedure introduces yet another set of resonances for which book-keeping is required. On the other hand, coupling to carbonyls is achieved with a minimal loss in sensitivity due to the large one-bond couplings between the carbonyls and the neighboring  $C_\alpha$  and N. On the whole, all these multidimensional through-bond experiments represent a powerful set of tools which permit assignments of almost all magnetic nuclei in proteins in a straightforward fashion.

### *3.2.3. Through-bond assignments in nucleic acids*

Recent advances in the biochemistry of nucleic acids have made it possible to introduce stable isotope labels in nucleic acids [90–92, 169,170]. This triggered the development of a series of new heteronuclear NMR experiments that are specifically geared towards nucleic acids [171–184]. In general, the assignment of NMR spectra of nucleic acids is considerably more difficult than corresponding assignments in proteins.

Proteins generally feature a wider range of chemical shift dispersion in part because proteins are composed of twenty building blocks, whereas nucleic acids are composed of only four nucleotides which differ only in their base moiety [50,187]. Ribonucleic acids (RNA) contain adenine, guanine, cytosine and uracil bases whereas in deoxy-ribonucleic acids (DNA) uracil is replaced by thymine. In addition, the base moieties of nucleotides possess only a small number of protons, some of which tend to be in rapid exchange with water, i.e. amino and imino protons. Excluding the exchange broadened amino protons, the base moieties contain only two to three observable protons whereas many amino acid sidechains contain a considerably larger number of observable protons. Finally, nucleic acids generally lack the tightly packed hydrophobic core [188] which is the hallmark of globular proteins, resulting in relatively poor chemical shift dispersion and scarcity of tertiary structural elements. On the other hand, there is a keen interest in gaining an understanding of how structural features influence the wealth of biochemical properties, particularly of RNA molecules.

A number of pulse sequences which were developed with proteins in mind can be readily applied to [ $^{15}\text{N}$ ,  $^{13}\text{C}$ ]-labeled nucleic acids [189–191]. For example, the HCCH–COSY [154] and TOCSY [155] techniques have been successfully applied to delineate connectivities in the sugar moieties of carbon labeled RNA oligonucleotides [191]. The HCCH–

TOCSY has also been employed to establish through-bond connectivities between H2 and H8 protons in adenine utilizing the extensive carbon-carbon connectivity network in this fused heterocycle [174,175].

It would be most desirable to design NMR experiments which permitted the visualization of through-bond connectivities between sequentially linked nucleotides. This would greatly facilitate sequential resonance assignments in nucleic acid as this has been shown to be the case in proteins. Nucleotides are linked by phosphate groups which carry the magnetically active spin-1/2 nuclei  $^{31}\text{P}$  (see Fig. 8). Direct three-bond couplings between the phosphorous and 5'/3' proton have been delineated and utilized in  $^{31}\text{P}$  observed HETCOR and proton observed HSQC experiments [185]. In favorable cases these experi-

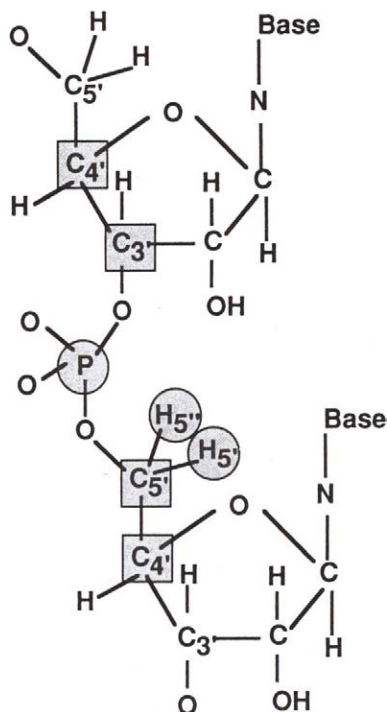


Fig. 8. Schematic representation of RNA backbone segment. Circled protons have non-zero  $3J_{\text{HP}}$  couplings to the depicted phosphorous atom. Squared carbon atoms feature non-zero two- and three bond  $J$ -couplings to the depicted phosphorous atom. These couplings can be utilized in hetero nuclear  $^1\text{H}$ - $^{31}\text{P}$  double resonance experiments [185], and  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{31}\text{P}$  triple resonance experiments [177,181,182,184,186].

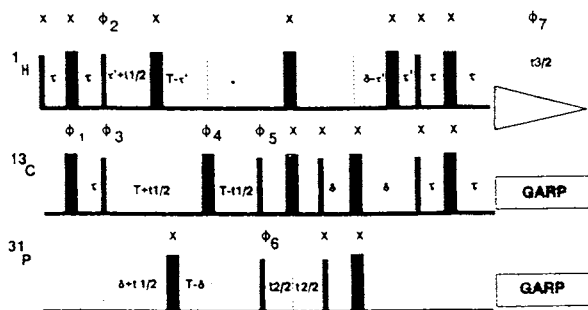


Fig. 9. Pulse sequence of HCP experiment [193]. The thick and thin bars represent  $180^\circ$  and  $90^\circ$  pulses, respectively. Constant time  $^{13}\text{C}$ -evolution is applied by moving the  $180^\circ$  pulse  $\Phi_4$  stepwise during the  $2T$  interval [193].  $\Phi_1 = x, -x$ ;  $\Phi_2 = y, -y$ ;  $\Phi_3 = x(+\text{TPPI}(t_1))$ ;  $\Phi_4 = 4x, 4y$ ;  $\Phi_5 = x, -x$ ;  $\Phi_6 = x, x, -x, -x (+\text{TPPI}(t_2))$ ;  $\Phi_7 = x, x, -x, -x, -x, x, x$  [177].

ments permit assignments of sequential connectivities depending on the  $^{31}\text{P}$ -chemical shift dispersion which tends to be modest [192]. Recently, Hilbers and coworkers developed  $^1\text{H}$ - $^{13}\text{C}$ - $^{31}\text{P}$  triple resonance HCP-experiment [193]. The pulse sequence of this HCP-experiment is depicted in Fig. 9. In this experiment  $\text{C4}'$  and  $\text{C5}'$  coherences evolve in  $t_1$ . A subsequent CP-INEPT transfer relays some of this carbon coherence onto the  $3'$  and  $5'$ -bonded  $^{31}\text{P}$  nuclei which evolve in  $t_2$ . Finally, this coherence is transferred back to  $\text{H4}'$  and  $\text{H5}'$  protons in two sequential reverse INEPT transfer steps. The feasibility of this experiment was demonstrated in a  $^{13}\text{C}$ -enriched RNA hairpin with the sequence  $5'(\text{GGGC}-\text{CAAA}-\text{GCCU})3'$ . Two-dimensional planes taken at the  $\text{C4}'$  and  $\text{C5}'$  resonances of the corresponding HCP-spectrum are depicted in Fig. 10. More recently several related experiments which establish through-bond correlations across the phosphodiester bond have been proposed [181,182,184,186].

Another important task in nucleic acids is the establishment of through-bond connectivities between the base and the carbohydrate moieties in nucleotides. Traditionally, these connectivities were determined by the 2D NOE experiment which relies on the spatial proximity

*Opposite:* Fig. 10. Sequentially arranged  $\text{C4}'$  and  $\text{C5}'$  planes from the HCP spectrum. The  $\text{C4}'$  planes show the sequential walk through the sugar-phosphate backbone from the first residue G1 (at the  $5'$ -end) to the final residue U12 (at the  $3'$ -end) as indicated by the arrows and the drawn lines. Adapted from Haus et al [177].



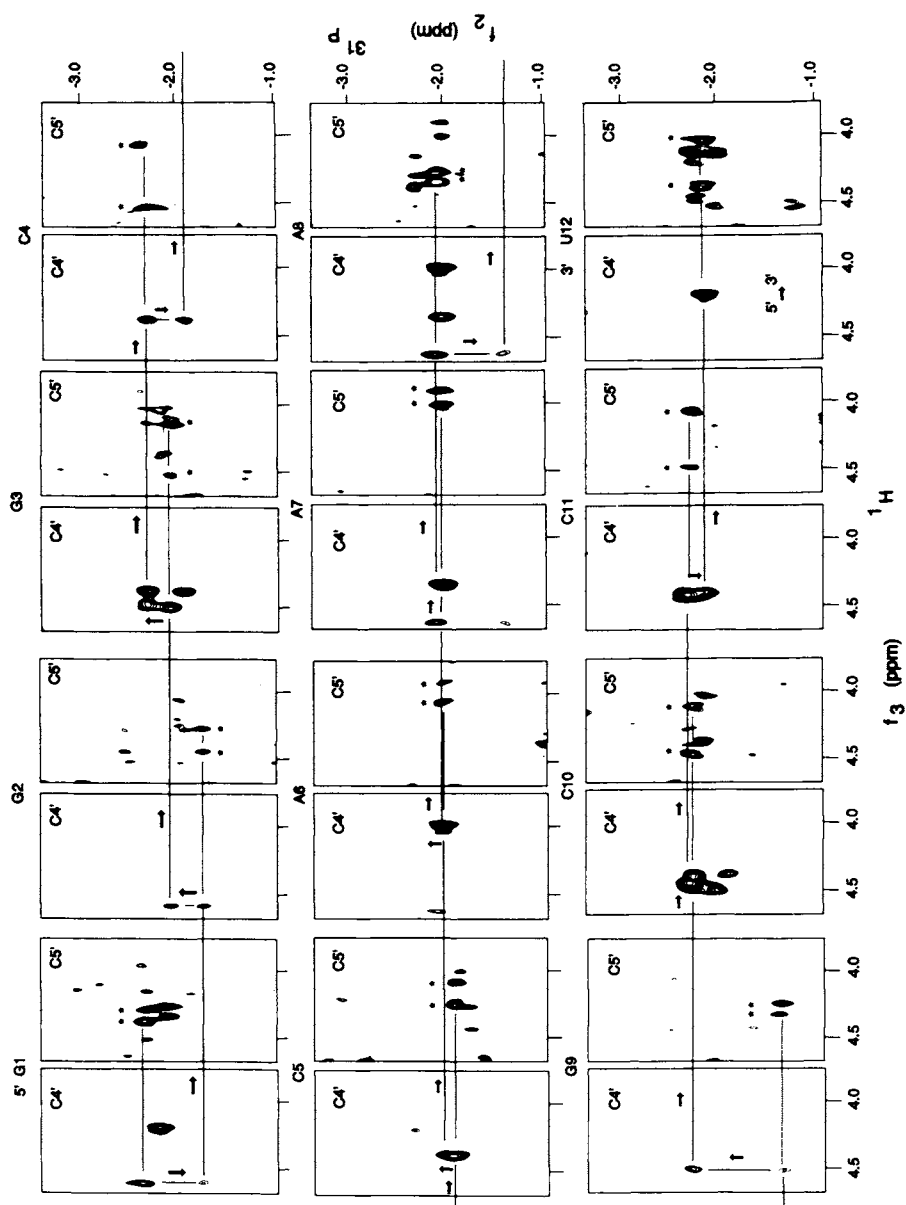


Fig. 10. Caption opposite

of anomeric protons to protons on the base predominantly H6 (pyrimidines) or H8 (purines) [50]. Not surprisingly, these through-space contacts may be fraught with ambiguity. Several heteronuclear double and triple resonance NMR methods have been developed recently that establish the desired through bond connectivities between anomeric protons and H6 protons in pyrimidines or H8 protons in purines respectively [172,173,176,178,180,194]. These experiments succeed in establishing through-bond magnetization transfers between anomeric protons and the H6 or H8 protons on the base utilizing one-bond  $^1J_{C1'-N1or9}$ ,  $^1J_{C6 \text{ or } 8-N1or9}$  couplings or  $^2J_{N9H8}$ -couplings. Considerable care had to be taken in these experiments to suppress undesirable  $J_{CN}$  and  $J_{CC}$ -couplings within the purine and pyrimidine rings.

### 3.3. Through-space correlations

Time-dependent dipolar couplings among protons is the primary source of relaxation for protons coupled to  $^{13}C$  and  $^{15}N$  nuclei in solution [195]. Individual protons (i) relax with the longitudinal relaxation rate  $\rho$  towards the Boltzmann equilibrium and exchange magnetization with neighboring protons (j) at the cross-relaxation rate  $\sigma_{ij}$ , where

$$\rho = 0.1\gamma^4\hbar^2 \sum (r_{ij}^{-6}) [6 J_2(\omega) + 3 J_1(\omega) + J_0(\omega)]$$

and

$$\sigma_{ij} = 0.1\gamma^4\hbar^2 (r_{ij}^{-6}) [6 J_2(\omega) - J_0(\omega)],$$

provided that fluctuations of inter-atomic distance between nuclei i and j can be neglected. These equations illustrate that cross-relaxation rates  $\sigma_{ij}$  between pairs of protons having an  $r_{ij}^{-6}$  dependence, and thereby inter-proton cross-relaxation is effective only among protons which are in close spatial contact. Hence, observation of a cross peak between a pair of protons in a NOESY spectrum severely restricts their possible spatial separation and thereby may serve as a convenient distance restraint. This restriction is both a curse and a blessing: long-range distance measurements are beyond the realm of solution NMR but interpretation of NOE contacts is greatly simplified by the steep distance dependence of inter-proton cross-relaxation. For historical reasons, NMR experiments which observe cross-relaxation effect [100] are called nuclear Overhauser spectroscopy or NOESY [135]. The basic

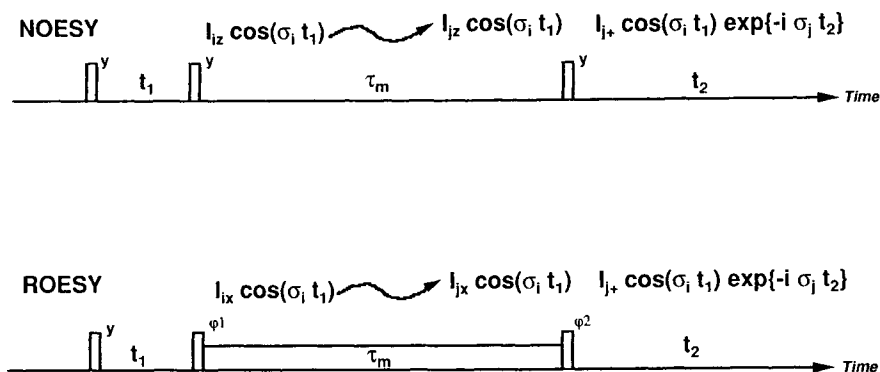


Fig. 11. Pulse sequences and relevant product operator terms of 2D NOESY and ROESY experiments.

scheme of the NOESY experiment [49] and its rotating frame analog (CAMEL SPIN [18] or ROESY [196]) is depicted in Fig. 11.

In the rotating frame NOE or ROESY experiment, which was invented by A. Bothner-By [18], the magnetization of interest is locked along a radio frequency (rf) field of constant frequency and amplitude such that the proton coherence of interest is locked along an effective Zeeman field which is tilted relative to the external magnetic field by an angle between  $45^\circ$  and  $135^\circ$ . The effect of this spin-locking is an altered cross-relaxation rate corresponding to an increase in the rate of transfer and an inversion of the sign of the transferred magnetization i.e. in the rotating frame the NOE is always positive (i.e. cross peaks are always  $180^\circ$  out of phase with diagonal peaks in the absence of isotropic mixing and chemical exchange). Hence, in a ROESY spectrum cross peaks between proton resonances always have opposite sign from the diagonal peak i.e. peaks arising from magnetization that did not exchange in the mixing period  $\tau_m$ . Hence, a comparison of a NOESY and a ROESY spectrum allows one to distinguish between cross-relaxation and a chemical exchange process where a pair of protons trade places giving rise to a cross peak that always is in phase with the diagonal peak [197].

NOESY/ROESY spectra tend to be far more complex than spectra arising from through-bond experiments [135,136]. First, the number of cross peaks in a NOESY spectrum of a macromolecule tends to be large. Second, cross peaks show a broad range of intensities. Hence, the number of observable cross peaks is determined by the sensitivity of the

spectrum and there is a considerable dynamic range in cross-peak intensities. Finally, the assignments of NOESY cross peaks is non-trivial: in the absence of any prior knowledge of a molecular structure almost any proton pair within a peptide chain could possibly give rise to an NOE cross peak. Even at very high magnetic fields multiple proton resonance pairs very often match the frequency coordinates of a typical NOE cross peak. Hence, the assignment process of NOESY spectra is laborious and fraught with pitfalls. For these reasons, extension of the NOESY spectrum into three and four dimensions dramatically reduces the overlap problem, and thereby greatly facilitates the assignments of cross peaks. The four-dimensional  $^{13}\text{C}$ - $^{13}\text{C}$  [4D CC-NOESY] and  $^{13}\text{C}$ - $^{15}\text{N}$ [4D CN-NOESY] spectra are particularly valuable since the resonances of both donor and acceptor proton are dispersed along an additional carbon-frequency axis. This additional dispersion very often suffices to uniquely define the proper assignment of an NOE cross peak.

The pulse sequences of two gradient enhanced 4D CC-NOESY experiments are depicted in Fig. 12 [198]. These experiments are straightforward extensions of the 2D NOESY [49,135] experiment. The carbon

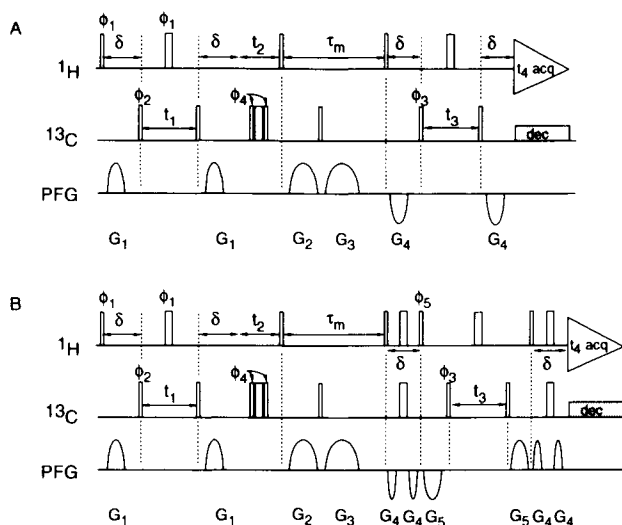


Fig. 12. Pulse schemes for the gradient versions of the 4D  $^{13}\text{C}/^{13}\text{C}$  separated HMQC-NOESY-HMQC (A) and HMQC-NOESY-HSQC (B) experiments [229].

frequencies are encoded as heteronuclear multiple quantum transitions (HMQC) [11] (Fig. 12a) or as carbon single quantum transitions (HSQC) (Fig. 12b) and evolve in  $t_1$  and  $t_3$ . The directly  $^{13}\text{C}$ -attached protons evolve in  $t_2$  and in the detection period  $t_4$ . This procedure assures that each proton resonance is associated with the frequency of its directly attached carbon. Consequently, assignment of NOE peaks is greatly simplified as illustrated in Fig. 13. In this figure two cross-sections through the 4D  $^{13}\text{C}/^{13}\text{C}$ -separated NOESY spectrum of IL-4 is depicted. These cross-sections display the  $^1\text{H}, ^{13}\text{C}$  correlations for all protons that have an NOE interactions with (A) Leu113- $\text{C}^{61}\text{H}_3$  and (B) Leu113- $\text{C}^{82}\text{H}_3$  or Lys88- $\text{C}_\gamma\text{H}_2$  [198].

Over the past few years several versions of  $^{15}\text{N}$  and  $^{13}\text{C}$  correlated 3D [28,36,199–208] and 4D [67,198,206,209–212] NOESY pulse sequences have been published. As in the sequences depicted in Fig. 12, all recent NOESY sequences that appeared in print after 1992 make extensive use of pulsed magnetic field gradients. These gradient pulses dephase unwanted coherences and provide considerably improved suppression of the universally unwanted water signal. This makes it feasible to record both carbon and nitrogen correlated NOESY spectra in a single sample in 90%  $\text{H}_2\text{O}$  resulting in a more precise alignment of resonances (frequently subtle differences in NMR resonance positions occur between  $\text{H}_2\text{O}$  dissolved and  $\text{D}_2\text{O}$  dissolved samples due to isotope induced changes in protonation state in carboxyl groups, as well as subtle differences in sample and buffer concentration). Furthermore, the pH reading of glass electrodes is affected by deuteration level, making a careful match of pH level more difficult [213]. Moreover, pulsed field gradients tend to simplify phase cycling schemes [214,215] considerably, and thereby permit major reductions in the minimum number of scans per time point. This makes it possible to collect many additional interferogram time points in a given time period. For example, in the 4D GE CC–NOESY experiment by Vuister et al. [198] the number of phase cycling steps has been reduced to 2 (Fig. 12a) or 4 (Fig. 12b) compared to 16 steps in the non-gradient version. This permits the collection of additional data points in the  $t_1$ ,  $t_2$  and  $t_3$  interferograms which eliminates the need for using linear prediction algorithm [96, 216–220] to artificially extend these interferograms prior to Fourier transformation. It is prudent to refrain from using these resolution enhancement techniques in NOESY-type spectra, because these computational algorithms do not work optimally in the presence of large variations in peak intensities or on low intensity peaks relative to the

spectral noise floor. The application of linear prediction and maximum entropy [221–228] becomes particularly precarious when the number of resonances along a given interferogram approaches the number of sampling points — a situation that is frequently encountered in 4D NOESY data of medium size proteins.

Pulse field gradients also make it practically feasible to devise clever pulse schemes which permit the simultaneous acquisition of  $^{15}\text{N}$ - and a  $^{13}\text{C}$ -correlated NOESY spectrum [206,230]. In terms of intrinsic resolution and ease of assignments the 4D NC- and CC-NOESY experiments are clearly preferred. Unfortunately, sensitivity is compromised in these experiments. First, each increase in dimensionality reduces the sensitivity by 30%. In addition, the necessary extra delays and rf pulses give rise to further attenuation in signal intensities due to  $T_2$ -relaxation and imperfections in the rf pulses. Finally, besides providing a convenient tool of increased spectral dispersion,  $^{13}\text{C}$  nuclei create additional fluctuating dipolar fields via the short-range one-bond  $^1\text{H}$ – $^{13}\text{C}$  interactions. This additional relaxation path causes further broadening of the proton lines. For these reasons, the overall sensitivity in a 4D CC-NOESY spectrum may be significantly lower than in a corresponding 2D proton NOESY spectrum in higher molecular weight proteins. Of course, one might argue that a comparison to a homonuclear 2D NOESY spectrum may not be reasonable in a large protein, where a quasi-continuum of cross peaks is encountered in most regions of the spectrum.

Fortunately, bacterial expression systems of proteins also permit the incorporation of deuterons into proteins [56,66,231–234]. Since deuterons have a low magnetic moment, dilution of protons by deuterons either by fractional deuteration or type specific deuteration may result in dramatic narrowing in line widths in the surviving protons. Equally important, dilution of protons reduces the cross-relaxation networks and thereby limits leakage of magnetization into spin-diffusion pathways [235]. Hence, in large proteins (>30 kDa) it might become neces-

---

*Opposite:* Fig. 13. (F1,F2) cross sections through the 4D  $^{13}\text{C}/^{13}\text{C}$ -separated 4D NOESY spectrum of IL-4, recorded with a 107 ms NOE mixing time. Cross sections shown are taken at the F3/F4 ( $^{13}\text{C}/^1\text{H}$ ) resonance frequencies of (A) Leu113– $\text{C}^{61}\text{H}_3$  and (B) the partially overlapping correlations of Leu113– $\text{C}^{62}\text{H}_3$  and Lys88– $\text{C}^7\text{H}_2$ . These cross sections display the  $^1\text{H}$ – $^{13}\text{C}$  correlations for all protons that have an NOE interaction with (A) Leu113– $\text{C}^{61}\text{H}_3$  and (B) Leu113– $\text{C}^{62}\text{H}_3$  or Lys88– $\text{C}^7\text{H}_2$ . Adapted from Ref. [229].

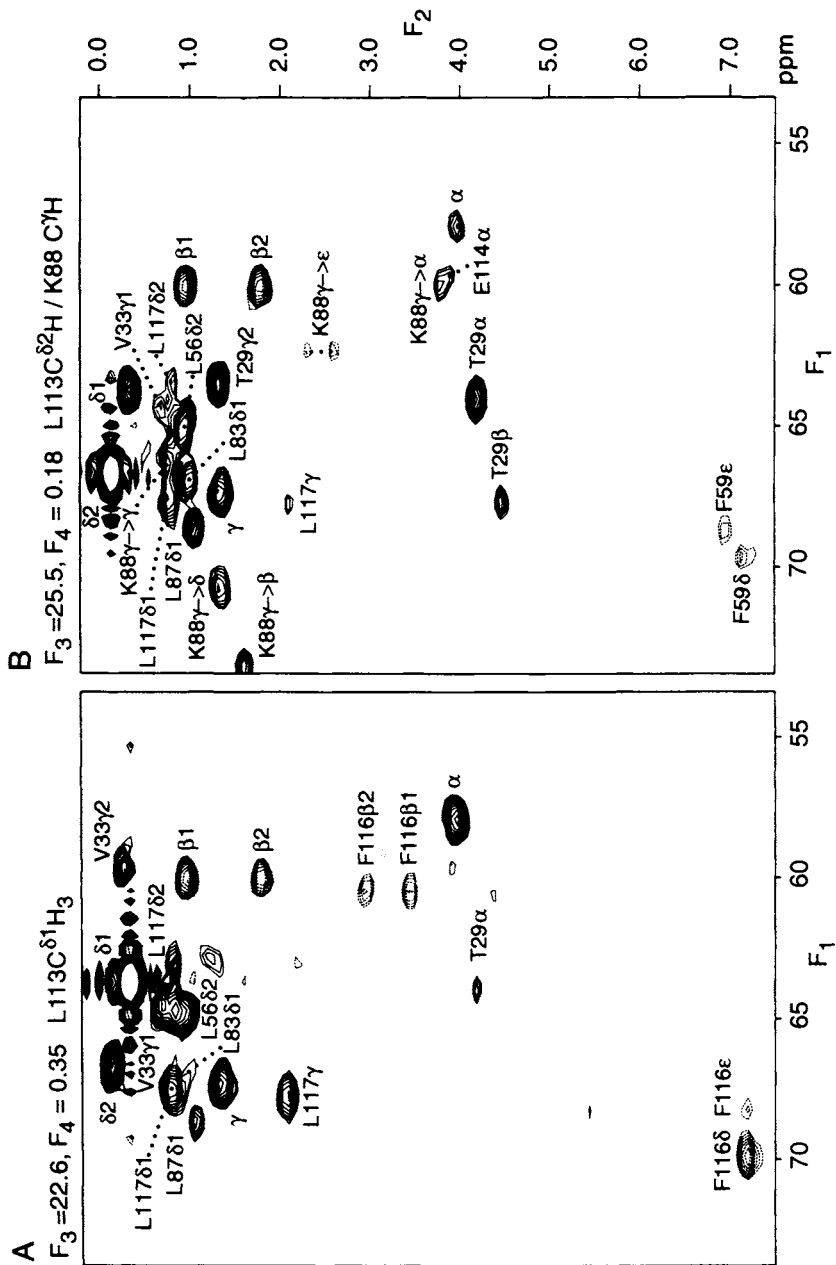


Fig. 13. Caption opposite.

sary to go back to 2D proton–proton NOESY experiments using samples of deuterated proteins in which a few types of amino acid residues are selectively protonated per sample. Such selectively protonated samples may yield sufficiently resolved and sensitive NOESY spectra but would require the preparation of a multitude of samples in order to elucidate a protein structure. Alternatively, randomly fractionally deuterated samples may yield a sufficiently sensitive NOESY spectra to afford the elucidation of medium resolution NMR structures of high molecular weight proteins.

## 4. ASSIGNMENT STRATEGIES

### 4.1. *Sequential resonance assignments*

In larger proteins it is advisable to focus resonance assignments on the peptide backbone first, and to pick up sidechain resonance assignments subsequently [146]. This strategy is aided by the ability to establish through-bond connectivities across peptide bonds using the heteronuclear multidimensional methods. For example, a sequential walk along the peptide backbone can be accomplished in a straightforward manner in 4D-HN(COCA)NH spectra [117] in which cross peaks directly connect  $^1\text{HN}$ ,  $^{15}\text{N}$  resonance pairs of residue (i) to their sequential  $^1\text{HN}$ ,  $^{15}\text{N}$  resonance pairs (i+1). This is illustrated in the four panels of Fig. 14 which depict a string of sequential NH-connectivities in deuterated [ $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-labeled calcineurin. However, spectra of lower dimensionality, in particular a combination of HNCACB, CBCA(CO)NH, HBHA(CO)NH and HNHAHB experiments, may provide sufficient information to perform a sequential walk along the peptide backbone. The HNCACB and CBCA(CO)NH experiments have the additional benefit of providing information about amino acid residue types via the  $\text{C}_\alpha$  and  $\text{C}_\beta$  carbon chemical shifts [121].

Several semi-automated resonance assignment programs have been developed which perform sequential resonance assignments in a partially automated fashion [236–252]. For example, the program by Friedrichs et al. divides the assignment process into three stages. In the first stage, groups of cross peaks which resonate at the same (NH,  $^{15}\text{N}$ ) coordinates are placed into separate bins or RIDs (Residue Information Data Structures) which represent generic amino acid residues. In the second stage, the program attempts to establish sequential strings of



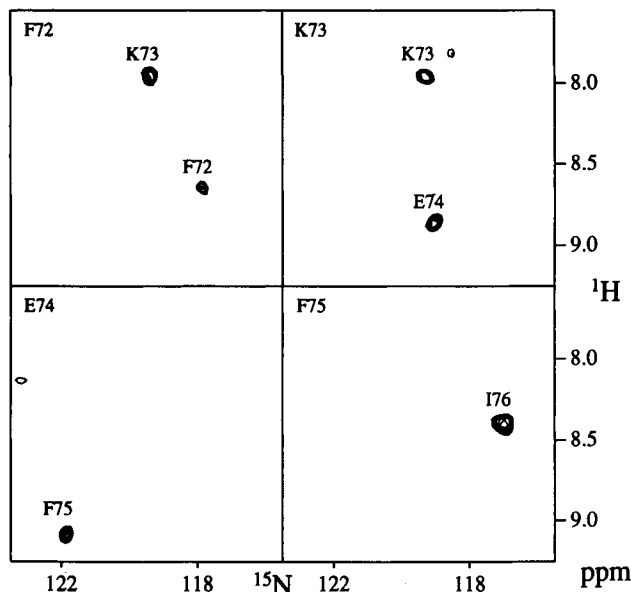


Fig. 14. Four ( $F_1$ ,  $F_2$ ) cross sections through the 4D HN(COCA)NH spectrum of calcineurin B (50%  $^2\text{H}_2\text{O}$ ), taken at the ( $F_3$ ,  $F_4$ ) frequencies of the amides of residues of Phe72–Phe75. Each cross section shows the connection to amide  $^{15}\text{N}$  and  $^1\text{H}$  frequencies of the sequential amino acid residue; panels for Phe72 and Lys73 also show weaker (4D) diagonal peaks to the same residue. Adapted from Ref. [117].

linkages among the RIDs. In stage 3A unique strings of RIDs are aligned with the amino acid sequence based on  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  chemical shift profiles utilizing the amino acid type characteristics of  $\text{C}_\alpha$  and  $\text{C}_\beta$  chemical shifts. In stage 3B gaps between unambiguously linked strings of RIDs are assigned by a clever search algorithm. In this approach it is of crucial importance that cross peaks are properly picked, and that artifactual humps and spikes are weeded out from the list of cross peaks.

Zimmerman et al. developed an automated assignment procedure which assigns both backbone and sidechain resonances [250]. Their procedure links resonances of HCCNH and HCC(CO)NH–TOCSY experiments [37,120,121,124]. This approach is particularly powerful in small to medium size proteins where NH-correlated carbon TOCSY can be collected with sufficient sensitivity. The number of publications on automated assignments keeps growing and level of sophistication in the

automated assignment procedure is continually increasing. Even commercial NMR data analysis programs have begun to feature computer assisted assignment algorithms.

#### **4.2. NOE assignments**

Unambiguous assignment of NOE cross peaks is considerably more difficult than performing sequential resonance assignments. In the absence of structural information almost any pair of protons could possibly give rise to an NOE contact. In light of this, it is prudent to be conservative in assigning NOE cross peaks to proton pairs because a few mis-assigned NOEs may give rise to improperly folded structures in which NOE violation may be distributed throughout the protein and thereby complicate the identification of improper NOE assignments.

It is desirable to underpin the spatial proximity of two sequentially well separated residues by a network of correlated NOE, as outlined in Fig. 15. This figure illustrates the identification of a spatial contact between a tyrosine and a threonine residue by two correlated NOEs ThrNH-TyrH<sub>c</sub> and Thrβ-TyrH<sub>δ</sub>. There, the two NOEs are separated on the threonine side by four chemical bonds and by three bonds on the tyrosine ring. Hence, these two NOEs have a correlation vector of {3,4}. Preferably, for each NOE contact there should be a correlated NOE whose correlation vector elements are small numbers. Recently, Kraulis [106] proposed a computer assisted procedure which permits automated assignments of 3D <sup>15</sup>N- and <sup>13</sup>C-correlated NOESY spectra without prior sequence specific NMR resonance assignments. This algorithm succeeds when nearly complete sets of NOE cross peaks can be obtained [106].

#### **4.3. Exclusively <sup>15</sup>N-labeled proteins**

It is not always economically feasible to incorporate <sup>13</sup>C labels in biological macromolecules, especially when the level of over-expression is low or when bacterial expression is not feasible. Due to the much lower cost of <sup>15</sup>N-labeled nutrients such as ammonium salts and <sup>15</sup>N-amino acids, it is possible to obtain sufficient amount of <sup>15</sup>N-expressed protein even if the yield is low or it may be possible to express <sup>15</sup>N-labeled proteins in CHO-cells by supplying the growth media with <sup>15</sup>N-labeled L-amino acids.

Any 2D proton NMR experiment which involves amide protons can be expanded into an <sup>15</sup>N-correlated three-dimensional analog in a

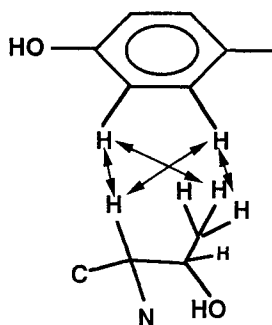


Fig. 15. Illustration of a hypothetical cluster of correlated long-range NOEs between a threonine residue and a tyrosine sidechain.

straightforward fashion either by temporarily transferring amide proton magnetization down to  $^{15}\text{N}$  single quantum transitions or by exciting proton–nitrogen zero- and double quantum transitions preferably immediately prior to detection of amide protons [31]. The most simple and important experiments in this class are the 3D  $^{15}\text{N}$ -edited proton HH–TOCSY and 3D  $^{15}\text{N}$ -edited NOESY experiments [28,253–255], and to a lesser extent the  $^{15}\text{N}$ -correlated COSY experiment [256]. The  $^{15}\text{N}$ -edited TOCSY experiments form the major tool for linking proton resonances within amino acid residues, featuring all the limitations of the homonuclear TOCSY i.e. inter-proton polarization transfers that are governed by relatively weak and conformation-dependent  $^3J_{\text{HH}}$ -coupling constants. The  $^{15}\text{N}$ -edited 3D NOESY spectra feature intra-residual, sequential and long-range NOEs. Cross peaks in NOESY spectra reflect spatial proximity regardless of the underlying chemical bonding network. In practice, sequential  $[\text{NH}, \text{H}_\alpha][\text{NH}, \text{NH}]$  and intra-residual NOEs tend to be the source of the most intense cross peaks.

Auxiliary methods to the straight 3D  $^{15}\text{N}$ -NOESY are the  $^{15}\text{N}$ -edited 3D TOCSY–NOESY and NOESY–TOCSY experiments [36]. In these experiments both a TOCSY and a NOESY transfer are concatenated back to back in the mixing period linking the evolution periods  $t_1$  and  $t_2$ . The main purpose of this approach is to increase the number of cross peaks between two adjacent amino acid residues and thereby increasing the chance of making an unambiguous assignment. In particular, the  $^{15}\text{N}$ -TOCSY–NOESY experiment has the benefit of producing a string of NH,NH cross peaks regardless of the relative arrangement of two sequential residues, which permits a straightforward walk along the peptide chain.

The 3D HNHB-experiment correlates  $H_\beta$ -protons directly to backbone amide nitrogens [257]. This is achieved by transferring amide proton magnetization down to the amide nitrogens after which the nitrogens are allowed to build up bilinear coupling terms to  $H_\beta$ -protons for a period  $\Delta$ . Subsequently, these terms are temporarily converted into  $H_\beta$ -magnetization which evolves in  $t_1$ . After encoding the  $H_\beta$ -frequencies a portion of the coherence is transferred back to amide protons which are detected in  $t_3$ .  $^{15}\text{N}$ -frequencies are encoded either in the transfer period to or from the  $H_\beta$ -protons. The intensities of cross peaks in the resulting HNHB-spectra are proportional to the  $^3J_{\text{N } H_\beta}$ -coupling constants. Hence, this technique provides both HNHB through-bond connectivities and estimates for  $^3J_{\text{N } H_\beta}$ -coupling constants which in conjunction with  $^3J_{\text{H}\alpha H_\beta}$  may provide stereo-specific assignments of  $H_\beta$ -protons. Finally, it is worth mentioning that there are few two-dimensional methods that permit the measurement of  $^3J_{\text{NH } H\alpha}$ -coupling constants, namely the J-modulated [ $^{15}\text{N}$ ,  $^1\text{H}$ ]-COSY [258] and the [ $^{15}\text{N}$ ,  $^1\text{H}$ ]-HMQC-J experiment [259] which are already described in the previous chapter on 2D NMR.

## 5. COUPLING CONSTANTS

Scalar or J spin-spin coupling constants represent an additional source of structural information in biological macromolecules. Particularly, J-couplings between vicinal pairs of magnetic nuclei relate to the torsion angle of a rotatable bond by way of the well-known Karplus equations [71–73, 260, 261]. In addition, one bond  $^1J_{\text{C}^\alpha\text{--H}^\alpha}$  couplings have been shown to possess  $\Phi$ ,  $\Psi$  angular dependence [129, 262]. In fact, the very first application of two dimensional NMR in proteins involved the measurement of homonuclear proton J-couplings using the now dormant 2D-J technique [263, 264]. In recent years, a considerable number of heteronuclear pulse sequences have been developed to measure a plethora of homonuclear [258, 265–287] and heteronuclear J-couplings [267, 274, 288–322].

There are basically two approaches for measuring spin-spin coupling constants by heteronuclear NMR techniques. The first class of experiments aims at generating E.COSY type cross peak pattern [265, 323]. There, the goal is to spread the two branches of a cross peak doublet by the desired vicinal J-coupling in one frequency domain and a large splitting frequently caused by a heteronuclear  $^1J_{\text{HX}}$ -coupling constant in a second frequency domain [267, 288]. The resulting large separation of

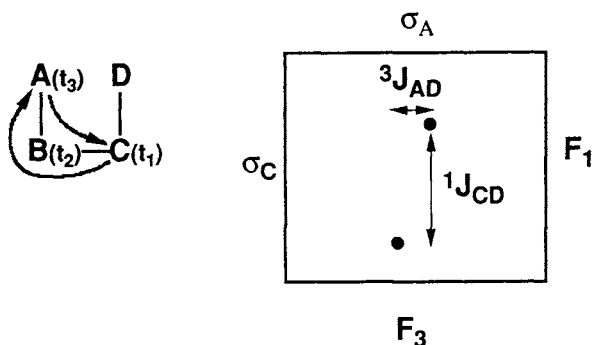


Fig. 16a. Measurement of J-coupling between the vicinal spin-pair A and D. The  ${}^3J_{AD}$ -coupling constant is encoded in the fine structure of the cross peak between A and the spy nucleus C. Coherence is transferred from A to C and is allowed to evolve during  $t_1$  under the influence of the one bond coupling  ${}^1J_{CD}$ . Subsequently, magnetization is transferred back to the nucleus A for detection during  $t_3$ . Additional dispersion can be achieved by evolution of the coherence of interest evolve during an additional indirect detection period  $t_2$  while it resides on spin B. The flow of magnetization is illustrated by the diagram on the left. The expected fine structure of the cross peak between the resonances of spin A and C is depicted in the panel on the right.

the peak doublet facilitates accurate detection of the respective centers of gravity, which then provides good estimates for the desired J-coupling. This is illustrated in Fig. 16a which depicts the general scheme of an E.COSY type experiment. First, magnetization is transferred to the spy nucleus C where it evolves in the period  $t_1$ . Then the magnetization is relayed back to A where its transverse magnetization is detected in  $t_3$  while spin C is decoupled. In the corresponding 2D spectrum the cross peak between A and C is split by the large  ${}^1J_{CD}$  in  $F_1$  and the desired vicinal  ${}^3J_{AD}$  in  $F_3$ . Resolution can be increased by evolving the magnetization along a optional third frequency axis while it resides on the intermediate nucleus B. By assigning A to NH, B to  ${}^{15}\text{N}$ , C to  ${}^{13}\text{C}_\alpha$  and D to  $\text{H}_\alpha$  one obtains the coherence path of the well-known HNCA-J experiment [269,272,283] which permits detection of  ${}^3J_{\text{H}_\text{N}\text{H}_\alpha}$ -coupling constants.

NMR wizard Ad Bax [257,293,324] devised alternate scheme of measuring J-coupling constants. Instead of trying to measure splittings within multiplets, the magnitudes of desired J-couplings are encoded in cross-peak intensities. One possible realization of this scheme is depicted in Fig. 16b.

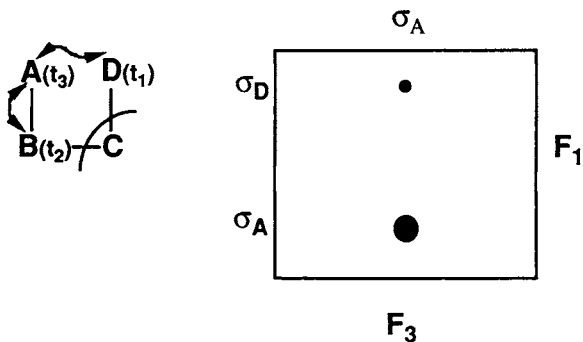


Fig. 16b. The  ${}^3J_{AD}$ -coupling constant is encoded in the intensity ratios of cross and diagonal peaks between spin A and D. The flow of magnetization in this class of experiments is depicted on the left. Magnetization is temporarily transferred from A to D by utilizing the  ${}^3J_{AD}$ -coupling. After evolving on D during  $t_1$ , the magnetization is transferred back to A where it is detected. Concurrent with this back transfer the magnetization is temporarily relayed to B utilizing the much larger one-bond coupling  ${}^1J_{AB}$ . The panel on the right depicts a hypothetical plane in a corresponding 3D spectrum taken at the resonance frequency of spin B. The intensity ratio between auto and cross peak encodes the size of the  ${}^3J_{AD}$ -coupling.

During a preparatory period, coupling terms between spin A and D develop which, in the product operator formalism, can be presented as

$$A_x \cos[\pi {}^3J_{AD} \tau] + A_y D_z \sin[\pi {}^3J_{AD} \tau].$$

A  $90^\circ$  pulse along the x-axis followed by the evolution period  $t_1$  converts this coherence into

$$A_x \cos[\pi {}^3J_{AD} \tau] \cos[2\pi \sigma_A t_1] + A_z D_y \sin[\pi {}^3J_{AD} \tau] \cos[2\pi \sigma_B t_1].$$

Termination of the  $t_1$ -evolution period by a  $90^\circ_x$ -pulse followed by a relay of magnetization back to A yields

$$A_x \cos[2\pi \sigma_A t_2] \{ \cos^2[\pi {}^3J_{AD} \tau] \cos[2\pi \sigma_A t_1] + \sin^2[\pi {}^3J_{AD} \tau] \cos[2\pi \sigma_B t_1] \}.$$

Hence, the ratios of diagonal to cross-peak intensities in the corresponding 2D spectrum are strongly dependent on the vicinal J-coupling between spin A and D. Here again, resolution is preferably enhanced by temporarily relaying all coherence of interest to spin B where it evolves

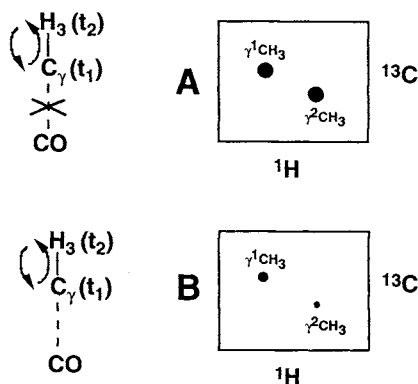


Fig. 16c. Measurement of  ${}^3J_{C\gamma CO}$ -coupling constants in valine methyls. Schematic representation of (A) A constant time proton-carbon HSQC spectrum with carbonyl decoupling and (B) The same spectrum without CO-decoupling. The ratio of cross-peak intensities are proportional to  $\cos(\pi {}^3J_{C\gamma CO} * \Delta)$ , where  $\Delta$  is the constant time evolution interval [293].

during an additional indirect evolution period  $t_2$ . The coherence pathway of the famous HNHA experiment [285] emerges by assigning A to NH, B to  ${}^{15}\text{N}$ , C to  $C_\alpha$  and D to  $H_\alpha$ . A second variation of this approach is depicted in Fig. 16c. Here, a two-dimensional  ${}^1\text{H}$ - ${}^{13}\text{C}/{}^{15}\text{N}$  spectrum is recorded with and without coupling of the  ${}^1\text{H}$ - ${}^{15}\text{N}$  to a third nuclei during  $t_1$ -evolution. Cross-peak intensities in corresponding difference spectra encode the magnitude of the  $J_{CX}$ -coupling constant. A good example of an experiment of this class is the  ${}^1\text{H}$ -detected long-range  ${}^{13}\text{C}$ - ${}^{13}\text{C}$  correlation experiment, which permits detection of vicinal  ${}^3J_{CC}$ -couplings between  $\gamma$ -methyls of valine and isoleucines and carbonyls [293].

A number of additional schemes of detecting J-coupling constants in two-dimensional spectra have been proposed whose discussion is beyond the scope of this chapter. Both the approaches discussed above have been utilized to measure homo-nuclear  ${}^3J_{HH}$ ,  ${}^3J_{C\gamma CO}$  constants and heteronuclear  ${}^3J_{NH}$ ,  ${}^3J_{CH}$  coupling constants.

### 5.1. Outlook on J-couplings

Most of the proposed detection schemes for measuring J-coupling constants work very well in small proteins whose overall motional correlation times are shorter than 10 ns. In larger proteins, however,

measurement of J-couplings involving protons are increasingly affected by proton dipolar relaxation [324–326]. Such distortions can be corrected by an explicit inclusion of the dipolar relaxation in the derivation of J-couplings from cross-peak intensities or J-splitting in the experiments mentioned above if the local geometric arrangements of all protons that contribute to the relaxation process is known [327,328]. Detection of J-couplings involving  $^{15}\text{N}$  and  $^{13}\text{C}$  tends to be limited in larger proteins by the small magnitude of many of these couplings. Short  $T_2$ -values or large line widths clearly limit the detection of these coupling constants. We anticipate that alternate detection schemes will be devised for measuring J-couplings in fully or fractionally deuterated proteins. For example, the distracting geminal dipolar relaxation could be easily suppressed by stereo-specific deuteration of methylene groups. Furthermore, the considerably extended  $T_2$  relaxation times in deuterated carbons may facilitate the detection of  $^3J_{\text{NC}}$  and  $^3J_{\text{CC}}$  coupling constants in higher molecular weight proteins.

## 6. DYNAMIC PROCESSES

### 6.1. $^{15}\text{N}$ relaxation

NMR relaxation phenomena are caused by fluctuating magnetic fields at the location of a magnetic nucleus. Even “static” structural information from NOE cross peaks is mediated by relaxation phenomena. Hence, most structural information one usually obtains in NMR experiments is based on proper interpretation of relaxation phenomena, primarily proton NOESY spectra. However, quantitative interpretation of proton relaxation processes is complicated by the large number of simultaneously interacting protons and by the presence of internal motions which cause both angular and radial fluctuations of dipolar couplings. Internal motions modulate both the angular and distance terms of inter-proton dipolar couplings and thereby complicate the extraction of accurate interproton distances from NOE cross-peak intensities. A detailed description of the dynamic studies of proteins is given in Chapter 5.

The relaxation of amide  $^{15}\text{N}$  nuclei is governed to a very good approximation by only two interactions, the dipolar coupling to its directly bonded amide proton and the nitrogen chemical shift tensor [329,330]. These two interactions solely depend on the orientational fluctuation of the NH-bond vector. Strict angular dependence considerably simplifies



interpretation of nitrogen relaxation parameters in the framework of the model-free approach [331,332]. Hence, these relaxation parameters are very suitable to compute estimates of amplitudes and time scales of internal motions. For these reasons,  $^{15}\text{N}$  measurements provide insight into internal flexibility of peptide backbone segments within a protein or nucleic acid (imino nitrogen). Pulse sequence virtuoso Lewis Kay and his coworkers developed sequences that permit accurate measurement of the following  $^{15}\text{N}$  relaxation parameters [329,333–335]: longitudinal relaxation time  $T_1$  which reflects the recovery time of Zeeman magnetization after an external perturbation, transverse relaxation time  $T_2$  which indicates the decay time of transverse magnetization and the  $^{15}\text{N}$  nuclear Overhauser enhancement factor (NOE) which represents the deviation of the equilibrium Zeeman polarization of a  $^{15}\text{N}$  nucleus when its directly bonded amide proton is being saturated by an rf irradiation. Several research groups besides Kay and coworkers proposed novel NMR experiments for indirect detection of  $^{15}\text{N}$  relaxation parameters [336] and suppression of cross-correlation in  $T_1$ ,  $T_2$  measurements [337–343]. The relaxation rates  $1/T_{1\text{N}}$ ,  $1/T_{2\text{N}}$  and the cross-relaxation rate  $\text{RN}(\text{HN} \rightarrow \text{Nz})$  is a function of spectral densities at the five frequencies: 0,  $\omega_{\text{N}}$ ,  $\omega_{\text{H}} - \omega_{\text{N}}$ ,  $\omega_{\text{H}}$  and  $\omega_{\text{H}} + \omega_{\text{N}}$ . Hence, one is faced with the problem of having intrinsically fewer experimental parameters than unknowns.

This problem has been addressed by most researchers by interpreting  $^{15}\text{N}$  relaxation data in the framework of a motional model. The most popular model proven to be the model-free approach by Lipari and Szabo [331,332]. The beauty of Szabo's approach is its ability to describe internal motions without having to know details about the exact internal motional dynamics. Instead, the model-free approach describes internal motions in terms of order parameters  $S$  which are related to overall amplitudes of motions and frequencies of internal motions. For example, an internal order parameter of a particular  $^{15}\text{N}$  nucleus of close to 1 implies only small amplitude internal fluctuations of the NH-bond vector. This simplified approach has been used extensively to delineate flexible segments within globular proteins [344–364]. Recently, Peng and Wagner [337,365] proposed the measurement of an additional three amide  $^{15}\text{N}$  and  $^1\text{H}$  relaxation parameters. This increases the number of experimental relaxation parameters to six, which permits a more comprehensive analysis of internal motions within the peptide backbone.

Ultra-slow motions, which occur in a time-scale that is considerably longer than the molecular tumbling rate, may give rise to the so called

exchange contributions to the transverse relaxation rate. These exchange contributions are proportional to the correlation times of these slow conformational rearrangements and the differences in the chemical shifts of the micro-environments that are experienced by a  $^{15}\text{N}$  spin during these motions. Recently, Wüthrich and coworkers [366] as well as Arseniev and coworkers [367] presented procedures for dissecting the exchange contributions in  $^{15}\text{N}$  transverse relaxation rates.

## **6.2. $^{13}\text{C}$ relaxation**

Motions within amino acid sidechains and/or nucleic acids can in principle be studied by  $^{13}\text{C}$  relaxation experiments. For example, Prendergast and coworkers [367] recently studied the motions of Trp28 and Trp31 in mutants of *E. coli* thioredoxin in which the two tryptophan residues were selectively  $^{13}\text{C}$ -labeled. Kay and co-workers [367] developed approaches for extracting carbon relaxation data from uniformly enriched protein samples, which as they pointed out is much more tricky both due to scalar  $J_{\text{CC}}$ -couplings and carbon-carbon dipolar relaxation paths.

## **6.3. Relaxation, chemical exchange with water**

Protons in water molecules can interact with protons on proteins or nucleic acids by either dipolar cross-relaxation or by chemical exchange. Both proteins and nucleic acids carry a considerable number of exchangeable protons. Exchangeable protons in proteins are located in amides, hydroxyls, indoles, protonated histidine sidechains, guanidinium- and amino groups in arginines and lysine respectively. In nucleic acids, protons in imino, amino and hydroxyl groups are exchangeable. Among the exchangeable protons, amide NH protons are most slowly exchanging. In general, the amide exchange rates reflect accessibility of an NH-group by water molecule. The hydrogen exchange itself is catalyzed by  $[\text{H}^+]$  and  $[\text{OH}^-]$  ions [102]. An exchangeable proton must get in direct contact with a water molecule before a hydrogen exchange can occur. Therefore, water molecules must be able to penetrate into a protein interior to exchange with a buried amide proton, suggesting that these exchange rates qualitatively sample conformational rearrangements within proteins and local unfolding events [102,368].

Exchange rates can be measured by monitoring the disappearance of NH-resonances after dissolving the protein sample of interest in a

D<sub>2</sub>O-buffer. Alternatively, rapid exchange rates and cross-relaxation originating from protein bound water molecules can be delineated by using a combination of water selective gradient enhanced 2D HMQC NOESY and ROESY experiments [369–376]. Besides identification of rapidly exchanging amide protons, these methods help to visualize bound water molecules. Many of these water molecules which can also be detected by X-ray diffraction are integral part of protein structures.

Under most circumstances, magnetization exchange between water and molecules of interest are more of nuisance than a blessing, because they tend to attenuate NH-signals, particularly at higher pH values (>6.5). This attenuation of amide proton magnetization is due to an unfavorable ratio of the T<sub>1</sub>-recovery time of bulk water relative to the typical relaxation delay of 1–2 seconds giving rise to a partial saturation of bulk water even if no lower CW-irradiation is utilized to suppress water signal. At sufficiently high pH values (pH > 6.5 and T > 25°C) the hydrogen exchange rate with water may exceed the non-selective T<sub>1</sub> recovery times for solvent exposed amide protons. This phenomenon propagates the saturation of bulk water onto amide protons, which may result in distinct reductions in signal intensities. Therefore, it is advisable to keep the bulk water fully polarized throughout a multidimensional NMR experiment [377–379]. Clever schemes have been devised recently in several heteronuclear experiments which induce a flip back of the water magnetization along the +z-axis at various stages of a pulse sequence in order to preserve the polarization of water [59,116,207,324,380–383]. Flip-back of H<sub>2</sub>O magnetization has been shown to result in considerable gains in sensitivity for solvent exposed protons. A simpler, as yet untried method would involve careful doping of the protein sample with a relaxation agent such as hyTempo [384,385] in order to reduce the T<sub>1</sub>-relaxation time of bulk water to 0.5–1 seconds. This would ensure proper recovery of NH-magnetization within the typical 1–1.5 seconds relaxation delay without adversely affecting inter-proton NOE cross-relaxation. However, it would be important to take precautions in choosing a relaxation agent that lacks binding affinity to the biological macromolecule of interest.

## 7. TOOLBOX OF PULSE SEQUENCE DESIGN

Pulse sequences of multidimensional time domain NMR experiments consist of strings of rf and gradient pulses interleaved with constant and/or variable intervals. Even the most awesome looking pulse sequence is

composed of simple functional blocks, which perform well-defined tasks such as, for example, transferring amide proton magnetization to amide nitrogens. During the course of designing an experiment, it is advisable to write down in detail the desired sequence of events. Next, one

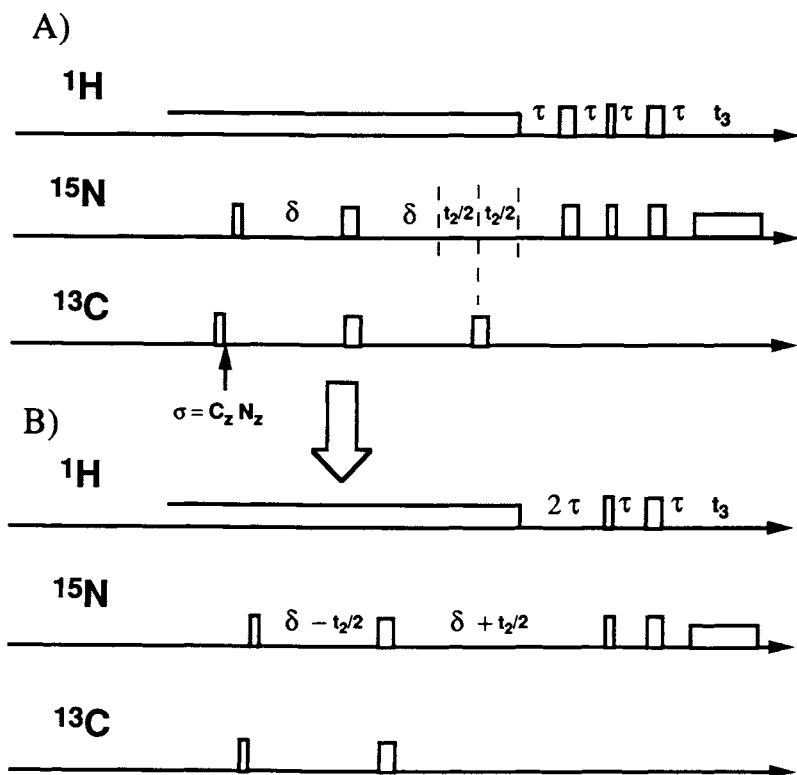


Fig. 17. Illustration of how synchronization of mutually non-interfering events may result in improved efficiency of a pulse sequence. A segment of the HNCACB experiment [112] which is responsible for relaying the  $C_\alpha/C_\beta$ -coherence after the evolution period  $t_1$  back to amide protons is shown. (A) Representation of a crude sequence in which only one event occurs at a time. During the  $\delta$ -period  $C_z N_y$ -coherence converts to  $N_x$  which then evolves under the influence of the  $^{15}\text{N}$ -chemical shift in  $t_2$ . Subsequently,  $N_x \cos(\sigma_N t_2)$  is transferred back to the amide proton during the  $\tau$ -period [133] (reverse INEPT). (B) In the optimized sequence these three events occur in parallel during the  $\delta$ -period, i.e. conversion of  $C_z N_y$  to  $N_x$ ,  $t_2$ -evolution of  $N_x$  and the conversion of  $N_x$  to  $N_y H_z$ . As a result of this, three  $180^\circ$  pulses become superfluous and the interval between the first and the last nitrogen  $90^\circ$  pulse is reduced by  $t_2 + 2\tau$ .

assembles known pulse modules in the proper sequence. Then one goes on implementing a phase cycling schemes or appropriately placed field gradient pulses which selects the desired coherence pathway.

While designing a phase cycling it is important to focus on suppressing the dominant and/or annoying artifacts, since it is generally beyond the realm of practicality to devise a complete phase cycling scheme that ensures full suppression of all unwanted coherence pathways. Once a crude sequence is assembled, it is a good idea to search for ways to optimize the pulse sequence. An important optimization strategy is to combine multiple events, where possible, and to let them unfold concurrently. An illustrative realization of this approach is shown in Fig. 17 which depicts a segment of the HNCACB-experiment [112]. The depicted pulse trains promote the back transfer of  $^{13}\text{C}$   $\text{C}_\gamma^\alpha \text{N}_z$  coherence to nitrogen, followed by  $^{15}\text{N}$   $t_2$  evolution and subsequent  $^{15}\text{N}$ - $^1\text{H}$  reverse INEPT transfer. Figure 17A depicts the raw sequence. In the optimized sequence all three events unfold concurrently (Fig. 17B), i.e.  $^{15}\text{N}$  unwind from the  $\text{C}_\alpha$ -carbons and recouple to the amide protons in the same  $\delta$ -interval, and time evolution of the  $^{15}\text{N}$  is implemented in constant time by stepping the  $^{15}\text{N}$ ,  $^{13}\text{C}$   $180^\circ$  pulse pair through the  $\delta$ -interval, thereby the nitrogen echo amplitudes are modulated by the  $^{15}\text{N}$  resonance frequencies. In the optimized sequence (Fig. 17B) there are only two  $180^\circ$  pulses remaining (versus five  $180^\circ$  pulses in Fig. 17A), and the duration of the sequence is shortened by the interval  $\{t_1 + 2 * \tau\}$ .

### 7.1. TOCSY versus INEPT/RELAY

There are basically two ways of shuffling magnetization by means of J-couplings among magnetic nuclei. First, magnetization can be transferred by invoking weak spin-spin couplings using the well-known relay of INEPT module [133], which consists of two delays flanking a  $180^\circ$  pulse (or pulse pair in a heteronuclear transfer) followed by a  $90^\circ$  pulse (or pulse pair in a heteronuclear transfer). Alternatively, the transfer can be performed in the rotating frame by subjecting magnetic nuclei to coherent rf irradiation with the aim of quenching the chemical shift Hamiltonian while preserving or enhancing spin-spin couplings. Under the resulting strong coupling conditions, magnetization can travel along a string of coupled nuclei. This rotating frame coherence transfer is generally called TOCSY [130,132] (total correlation spectroscopy) or Hartmann-Hahn polarization transfer (HOHAHA [19], HEHAHA [131,386-388]).

In homonuclear TOCSY transfers, the inter-nuclear couplings become twice as effective than those in the laboratory frame. For this reason, TOCSY is generally more effective than RELAY in homonuclear coherence transfer experiments although a RELAY transfer is generally more specific [389]. In heteronuclear transfers the benefits of TOCSY transfer are greatly diminished, because both INEPT and hetero TOCSY transfer occur at comparable rates. A clear cut sensitivity gain in hetero TOCSY experiments is assured only when resonances of interest are broadened by chemical exchange [390].

## **7.2. Phase cycling**

Besides pulsed field gradients and rf purging pulses, phase cycling of rf pulses represents the major tool for the selection of desired coherence pathways [214,215,391]. In order to select the desired coherence paths in a pulse sequence, the experiment is repeated, at a given  $t_1$ ,  $t_2$ ,  $t_3$ -time point,  $n$  times with the phase of the rf pulse being incremented by  $\Delta\phi$  in concert with a corresponding change in the receiver phase. The relationship between  $\Delta\phi$  of the rf pulse and the change in the receiver phase depends on the intended coherence level before and after the rf pulse. For example, if a  $90^\circ$  pulse converts single quantum transitions into Zeeman order, the coherence level changes by one and accordingly a  $180^\circ$  change of its rf phase causes an inversion of the detected signal in the receiver. Hence, in order to retain this path during the phase cycling the receiver phase must be inverted.

Most multidimensional NMR experiments contain a multitude of rf pulses which in many cases induce undesired changes in coherence levels in part due to imperfections in rf phase, rf field homogeneity or non-negligible resonance offsets. Obviously, it is impossible to cycle the phases of each individual pulse in a timely fashion. In practice, one tends to suppress only the major leaks in proper coherence paths and to tolerate minor spectral artifacts caused by improper coherence paths. Frequently, phase cycling is synchronized i.e. phases of two rf pulses move in concert (see for example Refs. [121,163]). This may lead to only partial but sufficient suppression of artifacts. Another neat trick is to move undesired coherences to the edge of a spectrum (States TPPI) [149]. This can be accomplished by alternating the phase of an excitation pulse with each  $t_1/t_2$ -increment with a concomitant inversion of the receiver phase. Any signal that is not affected by this pulse is automatically shifted by half the spectral width.

It is important to have at least one inversion of the receiver phase in a phase cycling scheme in order to remove dc-offset in the receiver signal. Quadrature image in the detection domain is removed by shuffling the coherence pathway, e.g. the x and y components of the magnetization are swapped in the acquisition domain during the phase cycling [392].

### **7.3. Timing**

It is important to precisely adjust pulse intervals. In particular, evolution delays must be precisely set such that the first point of an interferogram is sampled at  $t_x = 0$  or  $1/(2 \cdot SW_x)$  [150,393]. This ensures that the linear phase shift across a spectral domain is exactly 0 or  $180^\circ$ . This is essential for achieving flat baselines. It is also important to take into account linear phase shifts that are induced by selective rf pulses far outside the bandwidth of excitation [121,394]. Such undesirable phase distortions which are typically induced on coherences of aliphatic carbons by carbonyl decoupling pulses, can be eliminated either by the proper placement of additional selective pulses which induce a counter acting phase shift [121], or by the use of cosine modulated rf pulses which consist of two sidebands, one resonating at 170 ppm and the other at -90 ppm assuming that the carbon carrier is set to 40 ppm [152].

### **7.4. Gradients**

An excellent overview on this topic has been presented by Laue and coworkers [395]. Pulsed field gradient technology serves as a tool to make the external magnetic field very inhomogeneous across the NMR sample in a well-defined time period after which the high homogeneity is restored rapidly (preferably within less than one millisecond). During the gradient pulse, transverse coherences of the order greater than zero are rapidly dephased on a macroscopic level. This effect can be used to suppress unwanted coherence paths by temporarily flipping the coherences of interest back along the z-axis followed by a gradient pulse as depicted in Fig. 18a [49,396,397]. However, it is important to note that the dephased coherence is not lost in a real sense, because in the absence of significant lateral diffusion this coherence can be partially or fully restored by appropriate coherence transfer echoes, i.e. by the application of a gradient pulse of equal magnitude by the opposite polarity or by an echo pulse followed by an identical gradient pulse.

Therefore, if one wishes to suppress unwanted transverse coherences one has to be extremely careful in choosing an integrated proper gradient amplitude in a given gradient pulse, or else the suppressed coherence might form an unpleasant stimulated echo at a later time point within a pulse sequence. The emergence of multiple axis gradient probes will permit the application of a greater diversity of gradient pulses, and thereby will reduce the chance of unwanted refocussing of suppressed spin coherence. Another proven benefit of multiple axis gradients (gradient pulses along all three Cartesian coordinates) is the ability to apply gradients along an axis which is tilted by  $54.7^\circ$  (magic angle) relative to the external magnetic field. Such magic angle gradients effectively suppress effects caused by higher order polarization of bulk water [398–401]. These higher order polarization terms in the Boltzmann-equilibrium are negligible for protons in solutes, but these additional polarization terms show their nasty claws when the proton concentration approaches a whopping 100 molar (90%  $\text{H}_2\text{O}$ ). Hence, magic angle gradients may provide improved solvent peak suppression in aqueous solutions. However, single axis gradient probes may be fully adequate in experiments where the solvent equilibrium magnetization is fully restored prior to the acquisition period by means of water flip-back techniques. As mentioned before, a  $180^\circ$ -pulse sandwiched between two identical gradient pulses gives rise to a full restoration of the magnetization of the nuclei which are refocussed by the echo pulse, whereas coherences of all other nuclei are dephased (see Fig. 18b). Alternatively, if the two gradients flanking the  $180^\circ$ -pulse have equal magnitude but opposite signs, then coherence which is refocussed by

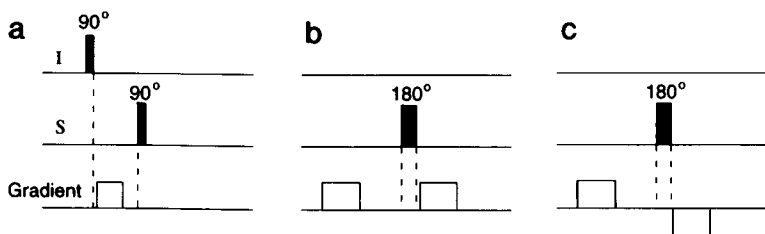


Fig. 18. Examples of different applications of pulsed field gradients in heteronuclear NMR [397]. (a) Selection of an  $I_z$  or  $I_z S_z$  intermediate; (b) selection of transverse S-spin magnetization which is being refocused by a  $180^\circ$  pulse; and (c) elimination of transverse S-spin components caused by an imperfect  $180^\circ$  decoupling pulse.



the 180°-pulse is selectively suppressed and all other coherences are retained (see Fig. 18c).

All the applications of field gradient pulses presented thus far focused on suppressing unwanted coherences. Alternatively, gradients can be used to exclusively retain a single coherence path while suppressing anything else. This can be achieved by the creation of a coherence transfer echo [402–408]. Representative sequences for generating coherence transfer echoes are depicted in Fig. 19. Application of a gradient G2 immediately after the  $^{15}\text{N}$  echo pulse in Fig. 19A dephases all transverse  $^{15}\text{N}$  coherences [409–411]. In the subsequent reverse INEPT transfer 50% of the dephased  $^{15}\text{N}$  coherence is transferred back to the amide protons. The application of a second gradient pulse G4 prior to the beginning of the acquisition period creates a  $^{15}\text{N}$ – $^1\text{H}$  coherence transfer echo, provided that the relative magnitudes of the two gradients G2 and G4 correspond to the gyromagnetic ratios of proton and  $^{15}\text{N}$  i.e. when G1 is about ten times stronger than G2. The refocused signal has the following dependence on the nitrogen  $t_1$  evolution period:

$$S_{\text{obs}}(G_1/G_2 = \gamma_{\text{H}}/\gamma_{^{15}\text{N}}) = 0.5 * H_x \cos(\sigma_{^{15}\text{N}} * t_1) - 0.5 * H_y * \sin(\sigma_{^{15}\text{N}} * t_1)$$

Hence, both x and y component are transferred with equal magnitude in a coherence transfer echo, which at the first glance seems to preclude the detection of phase-sensitive HSQC-spectra. Fortunately, by inverting the sign of gradient G1 the following altered coherence transfer echo is observed:

$$S_{\text{obs}}(G_1/G_2 = -1 * \gamma_{\text{H}}/\gamma_{^{15}\text{N}}) = 0.5 * H_x \cos(\sigma_{^{15}\text{N}} * t_1) + 0.5 * H_y * \sin(\sigma_{^{15}\text{N}} * t_1).$$

Hence, *post facto* formation of sums and differences of these two signals permits the reconstruction of the ordinary  $t_1$ -interferogram. As demonstrated by Kay and coworkers [408], it is possible to simultaneously collect both the coherence transfer echo and the anti-echo in a single scan and thereby doubling the sensitivity in one lucky strike by using the scheme depicted in Fig. 19B. This approach is closely related to the sensitivity enhancement method by Rance and co-workers [413]. The idea behind this clever scheme is to temporarily store the first reverse INEPT transferred component along the z-axis while the second component of x-nucleus magnetization is being back transferred to the corresponding proton. The effectiveness of this scheme is somewhat hampered when the x-nucleus is  $^{13}\text{C}$ , since the second component of

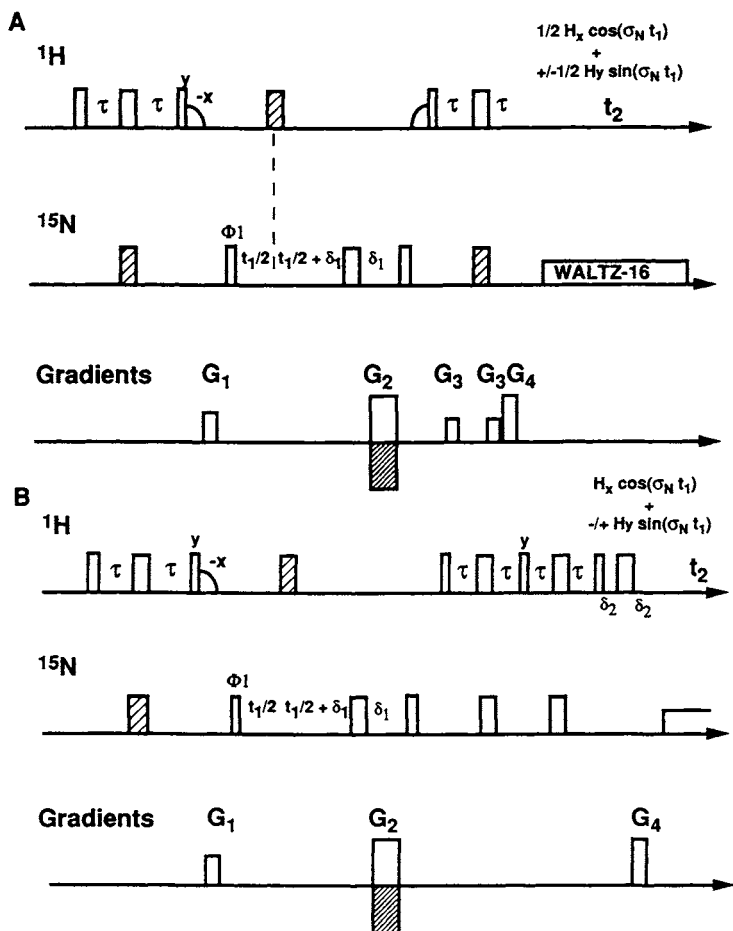


Fig. 19. Pulse sequence of gradient enhanced  $^1\text{H}^{15}\text{N}$ -HSQC experiment featuring selection of the desired coherence path by coherence transfer echoes. Narrow rectangles and wide rectangles represent  $90^\circ$  and  $180^\circ$  pulses, respectively. Hatched rectangles represent composite population inversion pulses. The quarter circles represent soft half Gaussian  $90^\circ$  pulses which ensure proper restoration of the water magnetization at the beginning of the acquisition time  $t_2$ . rf phases are  $x$ , except when marked otherwise,  $\Phi_1 = x, -x$ . The relative magnitudes of gradient pulses  $G_2$  and  $G_4$  are proportional to the nitrogen and proton gyromagnetic ratios. Each  $t_2$ -FID contains either the sum or the difference  $x$ - and  $y$  components of the  $^{15}\text{N}$   $t_1$ -interferogram depending on the relative sign of the  $G_2$ -gradient. For each  $t_1$ -point two FIDs are collected with  $G_2/G_4$  set to  $+$  or  $-$  ( $\gamma_{15\text{N}}/\gamma_{1\text{H}}$ ). (A) Straight coherence transfer echo experiment. (B) Sensitivity enhanced version, where both the echo and the anti echo are collected in a single FID [408,412].

carbon magnetization in  $\text{CH}_2$  and  $\text{CH}_3$ -groups is lost into multiple quantum coherence. Furthermore, during the additional INEPT delay some of the  $^{13}\text{C}$  coherence in methine groups is lost due to one-bond carbon-carbon couplings and carbon  $T_2$ -relaxation.

Overall, pulsed field gradients have become extremely useful tools in pulse sequence design. They dramatically improve solvent peak suppression, and allow the minimization of phase cycling steps thereby enabling the spectroscopist to collect a data set in either a *shorter* time period or to collect more sampling points along the indirectly detected interferograms.

## 8. HARDWARE REQUIREMENTS

Commercial high resolution NMR spectrometers have improved dramatically over the past decade in terms of sensitivity, stability and sophistication. First, stability and sensitivity/dynamic range are much more enhanced due to better probes, pulsed field gradient technology, low noise preamplifiers and high dynamic range receivers which have been further improved by digital filtering. Second, modern spectrometers can be equipped with up to four rf channels and multiple pulsed field gradient channels. Stability of the lock channel, homogeneities of magnetic fields and shim system are improved. Finally, the flexibility of pulse programmers in modern spectrometers permits the execution of very complex pulse sequences.

What is the best field strength for doing biological NMR studies? In general, both chemical shift dispersion and sensitivity increase with the magnetic field strength. However, one has to be aware of some complications. First, the line width contributions arising from chemical shift relaxation and chemical exchange processes increase with the magnetic field strength. Hence,  $^{19}\text{F}$  may give better resolved spectra at 400 MHz than at 750 MHz. Second, in high molecular weight biomolecules the proton longitudinal relaxation times tend to increase with increasing magnetic field strength and thereby counteracting somewhat the inherent sensitivity gains. The gain in raw sensitivity excluding relaxation effects, by increasing the proton carrier frequency from 600 MHz to 750 MHz is proportional to 47% based on the  $(\omega/2\pi)^{7/4}$  power law [414]. However, this gain in sensitivity has thus far only been realized in a proton observe single resonance probe head [415].

How many channels are needed? If money is no object, then four rf channels are most desirable. The fourth channel is essential for doing

$^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  (HCN) triple resonance experiments in deuterated proteins where the extra channel is needed for deuterium decoupling. In HCN triple resonance experiments the fourth channel is conveniently used for selectively pulsing carbonyl carbons. However, it is important to note that carbonyl pulses generated on the fourth channel can create artifacts on the aliphatic carbons if the rf phase of this channel is not correlated to the rf oscillator of the channel which pulses the aliphatic carbons. This can be avoided by synchronizing the rf phases of these two channels at the beginning of each scan. Alternatively, frequency shifted carbonyl pulses can be generated using the wave form generator of the channel which pulses carbons. Waveform generators permit amplitude and phase modulations of rf pulses. Particularly the rf channel which is used for pulsing carbons requires a waveform generator to create frequency shifted rf pulses and frequency-modulated ultra broadband decoupling schemes [416–418]

Pulsed field gradients have become indispensable in biological NMR both for reducing the number of phase cycling steps and for achieving efficient solvent peak suppression. Key quality criteria for pulsed field gradients are gradient strength, gradient recovery time and gradient pulse stability. The gradient field strength should exceed 30 G/cm (more is better as long as the probe can take it). In homonuclear proton experiments considerably larger gradients are needed to produce diffusion induced dephasing of the water signal [419]. Magnetic field homogeneity should fully recover within one millisecond after the application of a gradient pulse of 1–2 ms duration. Proper shaping (trapezoidal or sinusoidal) of gradient pulses reduces both recovery time and stress on the probe.

How many gradient axes are needed? Very good water suppression can already be achieved with a single axis NMR probe in which the gradient field is typically aligned along the z-axis (parallel to the external magnetic field). Additional gradients along the x- and/or y-axis provide additional flexibility. Simultaneous application of gradients along the x, y and z in a proper ratio permits the creation of the so-called magic angle gradient which works magic in suppressing higher order excitation states in bulk water, and thereby may result in additional suppression of water signal [398–400,420]. The magic angle gradient appears most effective in improving water signal suppression in homonuclear proton double quantum experiments [401]. In heteronuclear  $^1\text{H}$ – $^{13}\text{C}$  or  $^1\text{H}$ – $^{15}\text{N}$  experiments the usefulness of magic angle gradients is relatively limited [420]. Furthermore, water flip-back technique

greatly reduces multiquantum effects of bulk water [420]. At the very least, multiple axis gradients provide a larger diversity of dephasing schemes of unwanted signals, and thereby reducing the occurrence of undesirable stimulated echoes during the acquisition period.

## 9. OUTLOOK AND CONCLUDING REMARKS

*Deuteration:* Multidimensional NMR methods begin to fail when the intrinsic line widths approach or exceed the size of the utilized spin-spin couplings. For example,  $^{13}\text{C}$ – $^{13}\text{C}$  coherence transfer experiments which rely on the  $^1J_{\text{CC}}$ -couplings begin to fail in proteins whose molecular weight exceed 30 kDa. This is a fairly low molecular weight considering that many pharmacologically important proteins and RNA molecules exceed that molecular weight threshold. Does this mean that solution phase NMR is confined to dealing with wimpy little proteins whereas the crystallographers work on the more interesting systems?

Fortunately, it is possible to reduce the line widths in proteins by replacing protons with deuterons [231–233,235]. Close to full deuteration of proteins dramatically increases the  $T_2$  transverse relaxation times of carbons [421] by removing the major source of  $^{13}\text{C}$  relaxation — the one bond  $^{13}\text{C}$ – $^1\text{H}$  dipolar coupling. Furthermore, sizable line narrowing of amide protons is observed upon uniform deuteration of proteins [422]. Hence, the sensitivity of proton observed i.e. NH-correlated experiments is dramatically enhanced in deuterated proteins [59,117,125,423–426]. For this reason, sequential resonance assignments of amide protons and nitrogens and aliphatic carbons could be feasible in proteins up to a molecular weight of 55 kDa. At this molecular weight the amide nitrogen transverse relaxation time  $T_2(^{15}\text{N})$  may become the sensitivity limiting factor (B. T. Farmer II, personal communication).

The question then arises how to obtain NOE distance information in high molecular weight proteins. In fully protonated high molecular weight proteins (>30 kDa) the sensitivity of NOESY spectra becomes increasingly poor in part due to increasing line widths (low peak intensity) and excessive spin-diffusion which precludes substantial NOE build-up between most proton pairs. One way around this problem is to dilute protons by random fractional deuteration or type specific protonation of certain residues in an otherwise fully deuterated protein. It is expected that type specific protonation of key residues in active sites in, for example, enzymes while the remaining amino acid residues remain

deuterated, will play an important role in characterizing conformations of enzyme bound inhibitors. In such systems strong NOEs could possibly be observed between selectively protonated residues and the ligand, since a small number of protons would be surrounded by a wall of deuterons.

*Dimensionality:* Intrinsically, higher dimensional heteronuclear spectra provide in principle higher internal resolution at the expense of loss in sensitivity and a dramatic increase in instrument time for every additional new time domain, if one does not wish to excessively compromise the resolution in individual frequency domains. For most practical purposes four dimensions are the upper limit. In NOESY spectra four dimensions are ideal in terms of optimal frequency tagging of NOE cross peaks.

*Instrumentation:* The primary goals in NMR instrumentation are focused to gain further enhancements in sensitivity and improved spectral resolution. The remaining frontiers for sensitivity enhancements rely on devising higher field spectrometers and more efficient NMR probe heads. On the other hand, receivers in modern NMR spectrometer are already quite close to the theoretical optimum in terms of baseline noise and dynamic range. High field magnets offer both higher resolution and improved sensitivity although some of the intrinsic sensitivity gains are offset by increases in proton  $T_1$ -relaxation which perhaps could be compensated in part by clever use of relaxation agents.

Probe technology represents the key to extending the range of application of NMR spectroscopy. In current state-of-the-art NMR probes less than a tenth of one percent of the magnetic energy emitted by the precessing nuclei is picked up by the receiver coil. The primary challenge will be to enhance the magnetic coupling between the nuclear spins and the receiver coil while minimizing the effects of dielectric loss which is caused by the presence of electrolytes in the NMR resonator. Most recently, promising data have been presented on the use of high temperature superconducting materials in NMR probe design [427].

*Large volume probes:* A straightforward way of cranking up sensitivity is to increase the sample volume. However, even under ideal circumstances the NMR signal grows only in proportion to the square root of the sample volumes. Several factors counteract the full realization of this anticipated sensitivity gain: salt tolerance tends to decrease in larger volume samples, incomplete dissipation of heat during inverse decoupling may cause temperature gradients across the NMR sample, rf homogeneity and magnetic field homogeneity may become bigger challenges.

*Chemical shifts:* Chemical shifts are strongly dependent on local chemical environments, and consequently they encode structural information. Recently, Spera and Bax showed that secondary shifts of  $C_\alpha$  and  $C_\beta$  carbons tend to be correlated with secondary structures in proteins [428] through their dependence on  $\Phi$ ,  $\psi$  angles. Likewise,  $H_\alpha$  proton shifts have been correlated with helical and sheet structures in proteins (chemical shift index) [429,430]. These results indicate that the dominant contributions to  $^{13}\text{C}$ - and  $^{15}\text{N}$ - secondary chemical shifts could be correlated with local conformations with acceptable accuracy. In fact, chemical shift refinement terms have been incorporated in NMR structure calculation programs by several groups [431–436]. It has been shown that the judicious use of chemical shift restraints may improve the accuracy of 3D structures of proteins. It is important to note, however, that the correlations between chemical shifts and local conformations tend to be more complex than, for example, in J-couplings. While a vicinal J-coupling primarily depends on a single torsion angle, a  $C_\alpha$ -carbon shift is a function of at least two torsion angles ( $\Phi$ ,  $\psi$ ). For this reason, local conformational flexibility may lead to a complex averaging of chemical shifts. Hence, it is advisable to use chemical shift refinements of 3D structures with caution and only in regions of a biomolecule where high rigidity is indicated by a network of vicinal J-couplings and or local order parameters that were derived from relaxation data [436].

## REFERENCES

1. J. Jeener, Unpublished lecture. Ampere Summer School, Basko Polje, Yugoslavia, 1971.
2. L. Muller, A. Kumar and R.R. Ernst, *J. Chem. Phys.*, 63 (1975) 5490.
3. W.P. Aue, E. Bartholdi and R.R. Ernst, *J. Chem. Phys.*, 64 (1976) 2229.
4. R.R. Ernst, G. Bodenhausen and A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. The International Series of Monographs on Chemistry, Vol. 14. Clarendon Press, 1987.
5. A. Kumar, D. Welti and R.R. Ernst, *J. Magn. Reson.*, 18 (1975) 69.
6. G. Bodenhausen, R. Freeman, R. Niedermeyer and D.L. Turner, *J. Magn. Reson.*, 24 (1976) 291.
7. A.A. Maudsley and R.R. Ernst, *Chem. Phys. Lett.*, 50 (1977) 368.
8. G. Bodenhausen and R. Freeman, *J. Magn. Reson.*, 28 (1977) 471.
9. G. Drobny, A. Pines, S. Sinton, D. Weitekamp and D. Wemmer. *Fourier Transform Multiple Quantum NMR*, in *Faraday Symposium*. University of London, 1978.

10. A. Bax, A.F. Mehlkopf and J. Smidt, *J. Magn. Reson.*, 35 (1979) 167.
11. L. Mueller, *J. Am. Chem. Soc.*, 101 (1979) 4481.
12. G. Bodenhausen and D.J. Ruben, *Chem. Phys. Lett.*, 69 (1980) 185.
13. A. Kumar, R.R. Ernst and K. Wuthrich, *Biochem. Biophys. Res Commun.*, 95 (1980) 1.
14. M. Linder, A. Hoehener and R.R. Ernst, *J. Chem. Phys.*, 73 (1980) 4959.
15. G. Eich, G. Bodenhausen and R.R. Ernst, *J. Am. Chem. Soc.*, 104 (1982) 3731.
16. L. Braunschweiler and R.R. Ernst, *J. Magn. Reson.*, 53 (1983) 521.
17. M. Rance, O.W. Sørensen, G. Bodenhausen, G. Wagner, R.R. Ernst and K. Wüthrich, *Biochem. Biophys. Res. Commun.*, 117 (1983) 479.
18. A.A. Bothner-By, R.L. Stephens, J.-M. Lee, C.D. Warren and R.W. Jeanloz, *J. Am. Chem. Soc.*, 106 (1984) 811.
19. A. Bax and D.G. Davis, *J. Magn. Reson.*, 65 (1985) 355.
20. C. Griesinger, O.W. Sørensen and R.R. Ernst, *J. Am. Chem. Soc.*, 107 (1985) 6394.
21. M.A. Weiss, A.G. Redfield and R.H. Griffey, *Proc. Natl. Acad. Sci. USA*, 83 (1986) 1325.
22. C. Griesinger, G. Otting, K. Wuthrich and R.R. Ernst, *J. Am. Chem. Soc.*, 110 (1988) 7870.
23. C. Radloff and R.R. Ernst, *Mol. Phys.*, 66 (1989) 161.
24. A.M. Gronenborn, A. Bax, P.T. Wingfield and G.M. Clore, *FEBS Lett.*, 243 (1989) 93.
25. G. Bodenhausen and R.R. Ernst, *J. Magn. Reson.*, 45 (1981) 367.
26. C. Griesinger, O.W. Sørensen and R.R. Ernst, *J. Am. Chem. Soc.*, 109 (1987) 7227.
27. S.W. Fesik and E.R.P. Zuiderweg, *J. Magn. Reson.*, 78 (1988) 588.
28. D. Marion, L.E. Kay, S.W. Sparks, D.A. Torchia and A. Bax, *J. Am. Chem. Soc.*, 111 (1989) 1515.
29. C. Griesinger, O.W. Soerensen and R.R. Ernst, *J. Magn. Reson.*, 84 (1989) 14.
30. L.E. Kay, D. Marion and A. Bax, *J. Magn. Reson.*, 84 (1989) 72.
31. D. Marion, P.C. Driscoll, L.E. Kay, P.T. Wingfield, A. Bax, A.M. Gronenborn and G.M. Clore, *Biochemistry*, 28 (1989) 6150.
32. L.E. Kay, M. Ikura and A. Bax, *J. Am. Chem. Soc.*, 112 (1990) 888.
33. G.M. Clore and A.M. Gronenborn, *Progr. NMR Spectros.*, 23 (1991) 43.
34. E.R.P. Zuiderweg, A.M. Petros, S.W. Fesik and E.T. Olejniczak, *J. Am. Chem. Soc.*, 113 (1991) 370.
35. G.M. Clore, L.E. Kay, A. Bax and A.M. Gronenborn, *Biochemistry*, 30 (1991) 12.
36. L. Mueller, S. Campbell-Burk and P. Dommelle, *J. Magn. Reson.*, 96 (1992) 408.
37. T.M. Logan, E.T. Olejniczak, R.X. Xu and Fesik, *FEBS Lett.*, 314 (1992) 413.



38. S. Grzesiek and A. Bax, *J. Am. Chem. Soc.*, 114 (1992) 6291.
39. G.M. Clore and A.M. Gronenborn, in: *NMR of Proteins* (G.M. Clore and A.M. Gronenborn, eds.). CRC Press, Boca Raton, FL, 1993, pp. 1–307.
40. S. Grzesiek and A. Bax, *J. Biomol. NMR*, 3 (1993) 185.
41. B.A. Lyons, M. Tashiro, L. Cedergren, B. Bilsson and G.T. Montelione, *Biochemistry*, 32 (1993) 7839.
42. M. Wittekind and L. Mueller, *J. Magn. Reson., Ser. B*, 101 (1993) 201.
43. R.T. Clowes, W. Boucher, C.H. Hardman, P.J. Domaille and E.D. Laue, *J. Biomol. NMR*, 3 (1993) 349.
44. D.R. Muhandiram and L.E. Kay, *J. Magn. Reson., Ser. B*, 103 (1994) 203.
45. F.M. Ausubel, ed., *Current Protocols in Molecular Biology*. Wiley, New York, 1987.
46. J. Sambrook, E.F. Fritsch and T. Maniatis, *Molecular Cloning; A Laboratory Manual*. 2nd edn. Cold Spring Harbor Laboratory, 1989.
47. B. Lewin, *Genes* V. 1st edn. Oxford University Press and Cell Press, 1994.
48. K.B. Mullis and F. Ferre, *The Polymerase Chain Reaction*. 1st edn. Birkhauser, 1994.
49. J. Jeener, B.H. Meier, P. Bachmann and R.R. Ernst, *J. Chem. Phys.*, 71 (1979) 4546.
50. K. Wüthrich, *NMR of Proteins and Nucleic Acids*. Wiley, New York, 1986.
51. W. Braun, C. Bösch, L.R. Brown, N. Go and K. Wüthrich, *Biochim. Biophys. Acta*, 667 (1981) 377.
52. T.F. Havel and K. Wuthrich, *J. Mol. Biol.*, 182 (1985) 281.
53. G.M. Crippen, *J. Computational Physics*, 24 (1977) 96.
54. T.F. Havel, G.M. Crippen and I.D. Kuntz, *Biopolymers*, 18 (1979) 73.
55. N.J. Oppenheimer and T.L. James, ed. *Nuclear Magnetic Resonance Part B. Methods in Enzymology*, Vol. 177. 1989, pp. 7–49.
56. D.M. LeMaster, *Progr. NMR Spectros.*, 26 (1994) 371.
57. A. Bax and S. Grzesiek, *Acc. Chem. Res.*, 26 (1993) 131.
58. A. Bax, *Curr. Opin. Struct. Biol.*, 4 (1994) 738.
59. T. Yamazaki, W. Lee, C.H. Arrowsmith, D.R. Muhandiram and L.E. Kay, *J. Am. Chem. Soc.*, 116 (1994) 11655.
60. C.H. Arrowsmith, W. Lee, M. Revington, T. Yamazaki and L.E. Kay, *J. Am. Chem. Soc.*, 116 (1994) 6464.
61. M.L. Remerowski, T. Domke, A. Groenewegen and H.A.M. Pepermans, *J. Biomol. NMR*, 4 (1994) 257.
62. R.H. Fogh, D. Schipper, R. Boelens and R. Kaptein, *J. Biomol. NMR*, 5 (1995) 259.
63. D. Zhao, C.H. Arrowsmith, X. Jia and O. Jardetzky, *J. Mol. Biol.*, 229 (1993) 735.
64. P.J. Kraulis, P.J. Domaille, S.L. Campbell-Burk, T. Van Aken and E.D. Laue, *Biochemistry*, 33 (1994) 3515.
65. H. Zhang, D. Zhao, M. Revington, W. Lee, X. Jia, C. Arrowsmith and O.

- Jardetzky, J. Mol. Biol., 238 (1994) 592.
66. C.H. Arrowsmith, L. Treat-Clemons, L. Szilagyi, R. Pachter and O. Jardetzky, Makromol. Chem., Macromol. Symp., 34 (1990) 33.
  67. L.E. Kay, G.M. Clore, A. Bax and A.M. Gronenborn, Science, 249 (1990) 411.
  68. A.A. Maudsley, L. Mueller and R.R. Ernst, J. Magn. Reson., 28 (1977) 463.
  69. H. Oschkinat, C. Griesinger, P.J. Kraulis, O.W. Sørensen, R.R. Ernst, G.A. M. and G.M. Clore, Nature, 332 (1988) 374.
  70. G.W. Vuister, P. De Waard, R. Boelens, J.F.G. Vliegthart and R. Kaptein, J. Am. Chem. Soc., 111 (1989) 772.
  71. M. Karplus, J. Chem. Phys., 30 (1959) 11.
  72. M. Karplus, J. Am. Chem. Soc., 85 (1963) 2870.
  73. V.F. Bystrov, Progress in NMR Spectroscopy, 10 (1976) 41.
  74. G.W. Vuister, R. Boelens, A. Padilla, G.J. Kleywegt and R. Kaptein, Biochemistry, 29 (1990) 1829.
  75. J.P. Simorre, A. Caille, D. Marion, D. Marion and M. Ptak, Biochemistry, 30 (1991) 11600.
  76. X. Gao and W. Burkhardt, Biochemistry, 30 (1991) 7730.
  77. I. Radhakrishnan, D.J. Patel and X. Gao, Biochemistry, 31 (1992) 2514.
  78. S.W. Homans, Glycobiology, 2 (1992) 153.
  79. J. Habazettl, M. Schleicher, J. Otlewski and T.A. Holak, J. Mol. Biol., 228 (1992) 156.
  80. S. Bouaziz, C. Van Heijenoort, E. Guittet and J.Y. Lallemand, J. Chim. Phys. Phys.-Chim. Biol., 89 (1992) 147.
  81. R. Bernstein, A. Ross, C. Cieslar and T.A. Holak, J. Magn. Reson., Ser. B, 101 (1993) 185.
  82. D. Morikis, C.A. Lepre and P.E. Wright, Eur. J. Biochem., 219 (1994) 611.
  83. A. D'Ursi, H. Oschkinat, C. Cieslar, D. Picone, G. D'Alessio, P. Amodeo and P.A. Temussi, Eur. J. Biochem., 229 (1995) 494.
  84. A.P. Hansen, A.M. Petros, A.P. Mazar, T.M. Pederson, A. Rueter and S.W. Fesik, Biochemistry, 31 (1992) 12713.
  85. S.J. Archer, A. Bax, A.B. Roberts, M.B. Sporn, Y. Ogawa, K.A. Piez, J.A. Weatherbee, M.L.S. Tsang, R. Lucas, B.L., Zheng, J. Wenker and D.A. Torchia, Biochemistry, 32 (1993) 1152.
  86. S.J. Davis, M.J. Puklavec, D.A. Ashford, K. Harlos, E.Y. Jones, D.I. Stuart and A.F. Williams, Protein Eng., 6 (1993) 229.
  87. D.F. Wyss, J.M. Withka, M.H. Knoppers, K.A. Sterne, M.A. Recny and G. Wagner, Biochemistry, 32 (1993) 10995.
  88. K.L. Constantine, K.L. Colson, M. Wittekind, M.S. Friedrichs, N. Zein, J. Tuttle, D.R. Langley, J.E. Leet, D.R. Schroeder et al., Biochemistry, 33 (1994) 11438.
  89. C.G. Shibata, J.D. Gregory and B.S. Gerhardt, Arch. Biochem. Biophys., 319 (1995) 204.

90. E.P. Nikonowicz, A. Sirr, P. Legault, F.M. Jucker, L.M. Baer and A. Pardi, *Nucleic Acids Res.*, 20 (1992) 4507.
91. R.T. Batey, M. Inada, E. Kujawinski, J.D. Puglisi and J.R. Williamson, *Nucleic Acids Res.*, 20 (1992) 4515.
92. D.P. Zimmer and D.M. Crothers, *Proc. Natl. Acad. Sci. USA*, 92 (1995) 3091.
93. T. Kigawa, Y. Muto and S. Yokoyama, *J. Biomol. NMR*, 6 (1995) 129.
94. P.L. Weber and L. Mueller, *J. Magn. Reson.*, 81 (1989) 430.
95. E.R.P. Zuiderweg and S.W. Fesik, *Biochemistry*, 28 (1989) 2387.
96. G. Zhu and A. Bax, *J. Magn. Reson.*, 100 (1992) 202.
97. R.A. Hoffman and S. Forsen, *Progress in NMR*, 1 (1966) 15.
98. K. Wüthrich, *NMR in Biological Research: Peptides and Proteins*. Elsevier/North-Holland, Amsterdam, Oxford, New York, 1976.
99. J.H. Noggle and R.E. Schirmer, *The Nuclear Overhauser Effect*, Academic Press, New York, 1971.
100. A.W. Overhauser, *Phys. Rev.*, 92 (1953) 411.
101. S.W. Englander, N.W. Downer and H. Teitelbaum, *Annu. Rev. Biochem.*, 41 (1972) 903.
102. S.W. Englander and N.R. Kallenbach, *Q. Rev. Biophys.*, 16 (1983) 521.
103. R.R. Ernst, *Ber. Bunsen-Ges. Phys. Chem.*, 91 (1987) 1087.
104. A. Bax and S. Grzesiek, in: *NMR of Proteins* (M.G.M. Clore and A. Gronenborn, eds.). CRC, Boca Raton, 1993, pp. 33–52.
105. G.M. Clore and A.M. Gronenborn, in: *Methods in Enzymology* (T.L. James and N.J. Oppenheimer, eds.), Vol. 239. Academic Press, New York, 1994, pp. 349–363.
106. P.J. Kraulis, *J. Mol. Biol.*, 243 (1994) 696.
107. S. Grzesiek and A. Bax, *J. Magn. Reson.*, 99 (1992) 201.
108. S.L. Campbell-Burk, P.J. Dommelle, M.A. Starovasnik, W. Boucher and E.D. Laue, *J. Biomol. NMR*, 2 (1992) 639.
109. L.E. Kay, M. Wittekind, M.A. McCoy, M.S. Friedrichs and L. Mueller, *J. Magn. Reson.*, 98 (1992) 443.
110. E.T. Olejniczak, R.X. Xu, A.M. Petros and Fesik, *J. Magn. Reson.*, 100 (1992) 444.
111. W. Boucher, E.D. Laue, S.L. Campbell-Burk and P.J. Dommelle, *J. Biomol. NMR*, 2 (1992) 631.
112. M. Wittekind and L. Mueller, *J. Magn. Reson., Ser. B*, 101 (1993) 201.
113. R.T. Clubb, V. Thanabal and G. Wagner, *J. Biomol. NMR*, 2 (1992) 203.
114. S. Seip, J. Balbach and H. Kessler, *J. Magn. Reson.*, 100 (1992) 406.
115. M. Wittekind, W.J. Metzler and L. Mueller, *J. Magn. Reson., Ser. B*, 101 (1993) 214.
116. J. Stonehouse, R.T. Clowes, G.L. Shaw, J. Keeler and E.D. Laue, *J. Biomol. NMR*, 5 (1995) 226.
117. S. Grzesiek, J. Anglistter, H. Ren and A. Bax, *J. Am. Chem. Soc.*, 115 (1993) 4369.

118. M. Wittekind, W.J. Metzler and L. Mueller, *J. Magn. Reson., Ser. B*, 101 (1993) 214.
119. B.A. Lyons and G.T. Montellone, *J. Magn. Reson., Ser. B*, 101 (1993) 206.
120. G.T. Montelione, B.A. Lyons, S.D. Emerson and M. Tashiro, *J. Am. Chem. Soc.*, 114 (1992) 10974.
121. S. Grzesiek, J. Anglister and A. Bax, *J. Magn. Reson., Ser. B*, 101 (1993) 114.
122. E.T. Olejniczak, R.X. Xu, A.M. Petros and Fesik, *J. Magn. Reson.*, 100 (1992) 444.
123. W. Boucher, E.D. Laue, S. Campbell-Burk and P.J. Domaille, *J. Am. Chem. Soc.*, 114 (1992) 2262.
124. T.M. Logan, E.T. Olejniczak, R.X. Xu and Fesik, *J. Biomol. NMR*, 3 (1993) 225.
125. M. Shirakawa, M. Wälchli, M. Shimizu and Y. Kyogoku, *J. Biomol. NMR*, 5 (1995) 323.
126. O.W. Sorensen, G.W. Eich, M.H. Levitt, G. Bodenhausen and R.R. Ernst, *Progress in Nuclear Magnetic Resonance* (J.W. Emsley, J. Feeney and L.H. Sutcliffe, eds.), Vol. 16, Pergamon Press, Oxford, New York, 1983, pp. 163–192.
127. F. Delaglio, D.A. Torchia and A. Bax, *J. Biomol. NMR*, 1 (1991) 439.
128. G.W. Vuister, F. Delaglio and A. Bax, *J. Am. Chem. Soc.*, 114 (1992) 9674.
129. G.W. Vuister, F. Delaglio and A. Bax, *J. Biomol. NMR*, 3 (1993) 67.
130. S.R. Hartmann and E.L. Hahn, *Phys. Rev.*, 128 (1962) 2042.
131. L. Mueller and R.R. Ernst, *Mol. Phys.*, 38 (1979) 963.
132. L. Braunschweiler and R.R. Ernst, *J. Magn. Reson.*, 53 (1983) 521.
133. G.A. Morris and R. Freeman, *J. Am. Chem. Soc.*, 101 (1979) 760.
134. A. Bax, R. Freeman and T.A. Frenkiel, *J. Am. Chem. Soc.*, 103 (1981) 2102.
135. A. Kumar, R.R. Ernst and K. Wüthrich, *Biochem. Biophys. Res. Commun.*, 95 (1980) 1.
136. K. Nagayama, A. Kumar, K. Wüthrich and R.R. Ernst, *J. Magn. Reson.*, 40 (1980) 321.
137. T.F. Havel, G.M. Crippen, I.D. Kuntz and J.M. Blaney, *J. Theor. Biol.*, 104 (1983) 383.
138. D. Wemmer and N.R. Kallenbach, *Biochemistry*, 22 (1983) 1901.
139. E.R. Zuiderweg, R. Kaptein and K. Wuthrich, *Eur J Biochem*, 137 (1983) 279.
140. R.M. Scheek, R. Boelens, N. Russo, J.H. van Boom and R. Kaptein, *Biochemistry*, 23 (1984) 1371.
141. W. Braun and N. Go, *J. Mol. Biol.*, 186 (1985) 611.
142. G.M. Clore and A.M. Gronenborn, *FEBS Lett.*, 179 (1985) 187.
143. A.J. Wand and S.W. Englander, *Biochemistry*, 24 (1985) 5290.
144. R.B. Altman and O. Jardetzky, *J. Biochem.*, 100 (1986) 1403.
145. G.M. Clore, A.T. Bruenger, M. Karplus and A.M. Gronenborn, *J. Mol. Biol.*,

- 191 (1986) 523.
146. S.W. Englander and A.J. Wand, *Biochemistry*, 26 (1987) 5953.
147. L. Emsley and G. Bodenhausen, *Chem. Phys. Lett.*, 165 (1990) 469.
148. A.J. Shaka, J. Keeler, T. Frenkiel and R. Freeman, *J. Magn. Reson.*, 52 (1983) 335.
149. D. Marion, M. Ikura, R. Tschudin and A. Bax, *J. Magn. Reson.*, 85 (1989) 393.
150. A. Bax, M. Ikura, L.E. Kay and G. Zhu, *J. Magn. Reson.*, 91 (1991) 174.
151. M.A. McCoy and L. Mueller, *J. Am. Chem. Soc.*, 114 (1992) 2108.
152. M.A. McCoy and L. Mueller, *J. Magn. Reson.*, 98 (1992) 674.
153. G.M. Clore, A. Bax, P.C. Driscoll, P.T. Wingfield and A.M. Gronenborn, *Biochemistry*, 29 (1990) 8172.
154. L.E. Kay, M. Ikura and A. Bax, *J. Am. Chem. Soc.*, 112 (1990) 888.
155. A. Bax, G.M. Clore and A.M. Gronenborn, *J. Magn. Reson.*, 88 (1990) 425.
156. L.E. Kay, G.Y. Xu, A.U. Singer, D.R. Muhandiram and J.D. Forman-Kay, *J. Magn. Reson., Ser. B*, 101 (1993) 333.
157. E.T. Olejniczak, R.X. Xu and S.W. Fesik, *J. Biomol. NMR*, 2 (1992) 655.
158. M. Kadkhodaie, O. Rivas, M. Tan, A. Mohebbi and A. Shaka, *J. Magn. Reson.*, 91 (1991) 437.
159. T. Yamazaki, J.D. Forman-Kay and L.E. Kay, *J. Am. Chem. Soc.*, 115 (1993) 11054.
160. S. Grzesiek and A. Bax, *J. Am. Chem. Soc.*, 117 (1995) 6527.
161. L.E. Kay, M. Ikura, R. Tschudin and A. Bax, *J. Magn. Reson.*, 89 (1990) 496.
162. L.E. Kay, M. Ikura, G. Zhu and A. Bax, *J. Magn. Reson.*, 91 (1991) 422.
163. S. Grzesiek and A. Bax, *J. Magn. Reson.*, 96 (1992) 432.
164. S. Seip, J. Balbach and H. Kessler, *J. Biomol. NMR*, 3 (1993) 233.
165. S. Grzesiek and A. Bax, *J. Magn. Reson., Ser. B*, 102 (1993) 103.
166. L.E. Kay, *J. Magn. Reson., Ser. B*, 101 (1993) 110.
167. K. Dijkstra, G.J.A. Kroon and N.A.J. van Nuland, *J. Magn. Reson., Ser. A*, 107 (1994) 102.
168. B. Brutscher, F. Cordier, J.-P. Simorre, M. Caffrey and D. Marion, *J. Biomol. NMR*, 5 (1995) 202.
169. A. Ono, S. Tate, Y. Ishido and Kainosho, *J. Biomol. NMR*, 4 (1994) 581.
170. A. Ono, S. Tate and M. Kainosho, *Tanpakushitsu Kakusan Koso*, 40 (1995) 1509.
171. B.T. Farmer, II, L. Muller and E.P. Nikonowicz, *J. Am. Chem. Soc.*, 115 (1993) 11040.
172. V. Sklenar, M.R. Rejante, R.D. Peterson, E. Wang and J. Feigon, *J. Am. Chem. Soc.*, 115 (1993) 12181.
173. V. Sklenar, R.D. Peterson, M.R. Rejante and J. Feigon, *J. Biomol. NMR*, 3 (1993) 721.
174. P. Legault, B.T. Farmer, II, L. Mueller and A. Pardi, *J. Am. Chem. Soc.*, 116

- (1994) 2203.
175. J.P. Marino, J.H. Prestegard and D.M. Crother, *J. Am. Chem. Soc.*, 116 (1994) 2205.
176. V. Sklenar, R.D. Peterson, M.R. Rejante and J. Feigon, *J. Biomol. NMR*, 4 (1994) 117.
177. H.A. Heus, S.S. Wijmenga, F.J.M. van de Ven and C. Hilbers, *J. Am. Chem. Soc.*, 116 (1994) 4983.
178. B.T. Farmer II, L. Mueller, E.P. Nikonowicz and A. Pardi, *J. Biomol. NMR*, 4 (1994) 129.
179. G. Zhu, D. Live and A. Bax, *J. Am. Chem. Soc.*, 116 (1994) 8370.
180. S. Tate, A. Ono and M. Kainosho, *J. Am. Chem. Soc.*, 116 (1994) 5977.
181. S.S. Wijmenga, H.A. Heus, H.A.E. Leeuw, H. Hoppe, M. van der Graaf and C.W. Hilbers, *J. Biomol. NMR*, 5 (1995) 82.
182. J.P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D.M. Crothers and C. Griesinger, *J. Biomol. NMR*, 5 (1995) 87.
183. H. Schwalbe, J.P. Marino and S.J. Glaser, *J. Am. Chem. Soc.*, 117 (1995) 7251.
184. G. Varani, F. Aboul-ela, F. Allain and C.C. Gubser, *J. Biomol. NMR*, 5 (1995) 315.
185. G.W. Kellogg, A.A. Szewczak and P.B. Moore, *J. Am. Chem. Soc.*, 114 (1992) 2727.
186. S.-I. Tate, A. Ono and M. Kainosho, *J. Magn. Reson., Ser. B*, 106 (1995) 89.
187. K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Elsevier, New York, 1976.
188. T.E. Creighton, *Proteins, Structures and Molecular Principles*. Freeman, New York, 1984.
189. E.P. Nikonowicz and A. Pardi, *Nature*, 355 (1992) 184.
190. E.P. Nikonowicz and A. Pardi, *J. Am. Chem. Soc.*, 114 (1992) 1082.
191. E.P. Nikonowicz and A. Pardi, *J. Mol. Biol.*, 232 (1993) 1141.
192. P. Legault and A. Pardi, *J. Magn. Reson., Ser. B*, 103 (1994) 82.
193. S.S. Wijmenga, H.A. Heus, F.J.M. van de Ven and C. Hilbers, *NATO ASI Ser., Ser. H*, 87 (1994) 307.
194. B.T. Farmer, II, L. Muller, E.P. Nikonowicz and A. Pardi, *J. Am. Chem. Soc.*, 115 (1993) 11040.
195. A. Abragam, *The Principles of Nuclear Magnetism*. The International Series of Monographs on Physics (R.K. Adair and S.F. Edwards, eds.), Clarendon Press, Oxford, 1961.
196. A. Bax and D.G. Davis, *J. Magn. Reson.*, 63 (1985) 207.
197. D.G. Davis and A. Bax, *J. Magn. Reson.*, 64 (1985) 533.
198. G.W. Vuister, G.M. Clore, A.M. Gronenborn, R. Powers, D.S. Garrett, R. Tschudin and A. Bax, *J. Magn. Reson., Ser. B*, 101 (1993) 210.
199. M. Ikura, L.E. Kay, R. Tschudin and A. Bax, *J. Magn. Reson.*, 86 (1990) 204.

200. T. Frenkiel, C. Bauer, M.D. Carr, B. Birdsall and J. Feeney, *J. Magn. Reson.*, 90 (1990) 420.
201. C. Dalvit, G. Bovermann, K. Memmert and M. Zurini, *J. Magn. Reson.*, 96 (1992) 174.
202. K.A. Carpenter and F. Ni, *J. Magn. Reson.*, 99 (1992) 192.
203. G.W. Vuister, R. Boelens, R. Kaptein, M. Burgering and P.C.M. Van Zijl, *J. Biomol. NMR*, 2 (1992) 301.
204. A. Majumdar and E.R.P. Zuiderweg, *J. Magn. Reson., Ser. B*, 102 (1993) 242.
205. D.R. Muhandiram, N.A. Farrow, G.Y. Xu, S.H. Smallcombe and L.E. Kay, *J. Magn. Reson., Ser. B*, 102 (1993) 317.
206. S.M. Pascal, D.R. Muhandiram, T. Yamazaki and J.D. Forman-Kay, *J. Magn. Reson., Ser. B*, 103 (1994) 197.
207. W. Jahnke, M. Baur, G. Gemmecker and Kessler, *J. Magn. Reson., Ser. B*, 106 (1995) 86.
208. J.H. Davis, *J. Biomol. NMR*, 5 (1995) 433.
209. E.R.P. Zuiderweg, A.M. Petros, S.W. Fesik and E.T. Olejniczak, *J. Am. Chem. Soc.*, 113 (1991) 370.
210. G.M. Clore, L.E. Kay, A. Bax and A.M. Gronenborn, *Biochemistry*, 30 (1991) 12.
211. D.R. Muhandiram, G.Y. Xu and L.E. Kay, *J. Biomol. NMR*, 3 (1993) 463.
212. B.T. Farmer and L. Mueller, *J. Biomol. NMR*, 4 (1994) 673.
213. P.K. Glasoe and F.A. Long, *J Phys Chem*, 64 (1960) 188.
214. G. Bodenhausen, H. Kogler and R.R. Ernst, *J. Magn. Reson.*, 58 (1984) 370.
215. A.D. Bain, *J. Magn. Reson.*, 56 (1984) 41.
216. D.S. Stephenson, *Progress in NMR Spectroscopy* 20 (1988) 515.
217. J.C. Hoch, *Methods Enzymol.*, 176 (1989) 216.
218. E.T. Olejniczak and H.L. Eaton, *J. Magn. Reson.*, 87 (1990) 628.
219. G. Zhu and A. Bax, *J. Magn. Reson.*, 90 (1990) 405.
220. P. Koehl, C. Ling and J.F. Lefevre, *J. Magn. Reson., Ser. A*, 109 (1994) 32.
221. E.D. Laue, J. Skilling, J. Staunton, S. Sibisi and R.G. Brereton, *J. Magn. Reson.*, 62 (1985) 43.
222. E.D. Laue, M.R. Mayger, J. Skilling and J. Staunton, *J. Magn. Reson.*, 68 (1986) 14.
223. M.A. Delsuc, M. Robin, C. Van Heijenoort, C.B. Reisdorf, E. Guittet and J.Y. Lallemand, *NATO ASI Ser., Ser. A*, 225 (1991) 163.
224. J. Skilling, *Am. Lab.*, 24(15) (1992) 32J-32M.
225. P. Schmieder, A.S. Stern, G. Wagner and J.C. Hoch, *J. Biomol. NMR*, 3 (1993) 569.
226. G.W. Jeong, P.N. Borer and S.S. Wang, *J. Magn. Reson., Ser. A*, 103 (1993) 123.
227. P. Schmieder, A.S. Stern, G. Wagner and Hoch, *J. Biomol. NMR*, 4 (1994) 483.

- 228. S. Wang, I. Pelczer and P.N. Borer, *J. Magn. Reson., Ser. A*, 108 (1994) 171
- 229. G.W. Vuister and A. Bax, *J. Magn. Reson., Ser. B*, 101 (1993) 210
- 230. B.T. Farmer and L. Mueller, *J. Biomol. NMR*, 4 (1994) 673
- 231. D.M. LeMaster and F.M. Richards, *Biochemistry*, 27 (1988) 142.
- 232. D.A. Torchia, S.W. Sparks and A. Bax, *J. Am. Chem. Soc.*, 110 (1988) 2320.
- 233. D.M. LeMaster, *Methods Enzymol.*, 177 (1989) 23.
- 234. D.M. LeMaster, *Annu. Rev. Biophys. Biophys. Chem.*, 19 (1990) 243.
- 235. D.M. LeMaster, *FEBS Lett.*, 233 (1988) 326.
- 236. M. Billeter, V.J. Basus and I.D. Kuntz, *J. Magn. Reson.*, 76 (1988) 400.
- 237. C. Cieslar, G.M. Clore and A.M. Gronenborn, *J. Magn. Reson.*, 80 (1988) 119.
- 238. P.J. Kraulis, *J. Magn. Reson.*, 84 (1989) 627.
- 239. A.J. Wand and S.J. Nelson, *Trans. Am. Crystallogr. Assoc.*, 24 (1988) 131.
- 240. P.L. Weber, A. Malikayil J and L. Mueller, *J. Magn. Reson.*, 82 (1989) 419.
- 241. P. Catasti, E. Carrara and C. Nicolini, *J. Computational Chemistry*, 11 (1990) 805.
- 242. F.J.M. Van de Ven, P.O. Lycksell and A. Van Kammen, *Eur. J. Biochem.*, 190 (1990) 583.
- 243. J.C. Hoch, C. Redfield and A.S. Stern, *Curr. Opin. Struct. Biol.*, 1 (1991) 1036.
- 244. R. Wehrens, C. Lucasius and L.K. Buydens, *J. Chem. Inf. Comput. Sci.*, 33 (1993) 245.
- 245. R. Bernstein, C. Cieslar, A. Ross, H. Oschkinat, J. Freund and T.A. Holak, *J. Biomol. NMR*, 3 (1993) 245.
- 246. G.J. Kleywegt, G.W. Vuister, A. Padilla, R.M.A. Knegt, R. Boelens and R. Kaptein, *J. Magn. Reson., Ser. B*, 102 (1993) 166.
- 247. G.V.T. Swapna and R. Ramachandran, *J. Magn. Reson.*, B102 (1993) 204.
- 248. M.S. Friedrichs, L. Mueller and M. Wittekind, *J. Biomol. NMR*, 4 (1994) 703.
- 249. J.B. Olson, Jr. and J.L. Markley, *J. Biomol. NMR*, 4 (1994) 385.
- 250. D. Zimmerman, C. Kulikowski, L. Wang and B. Lyons, *J. Biomol. NMR*, 4 (1994) 241.
- 251. B.J. Hare and J.H. Prestegard, *J. Biomol. NMR*, 4 (1994) 35.
- 252. N. Morelle, B. Brutscher and J.-P. Simorre, *J. Biomol. NMR*, 5 (1995) 154.
- 253. D. Marion, P.C. Driscoll, L.E. Kay, P.T. Wingfield, A. Bax, A.M. Gronenborn and G.M. Clore, *Biochemistry*, 28 (1989) 6150.
- 254. A.M. Gronenborn, A. Bax, P.T. Wingfield and G. Clore, *FEBS Lett.*, 243 (1989) 93.
- 255. A.M. Gronenborn, P.T. Wingfield and G.M. Clore, *Biochemistry*, 28 (1989) 5081.
- 256. G.M. Clore, A. Bax, P. Wingfield and A.M. Gronenborn, *FEBS Lett.*, 238 (1988) 17.
- 257. S.J. Archer, M. Ikura, D.A. Torchia and A. Bax, *J. Magn Reson.* 95 (1991) 636.



- 258. D. Neri, G. Otting and K. Wuethrich, *J. Am. Chem. Soc.*, 112 (1990) 3663.
- 259. L.E. Kay and A. Bax, *J. Magn. Reson.*, 86 (1990) 110.
- 260. K.D. Kopple, G.R. Wiley and R. Tauke, *Biopolymers*, 12 (1973) 627.
- 261. A. Pardi, M. Billeter and K. Wüthrich, *J. Mol. Biol.*, 180 (1984) 741.
- 262. D.F. Mierke, S.G. Grdadolnik and H. Kessler, *J. Am. Chem. Soc.*, 114 (1992) 8283.
- 263. W.P. Aue, J. Karhan and R.R. Ernst, *J. Chem. Phys.*, 64 (1976) 4226.
- 264. K. Nagayama, K. Wuethrich, P. Bachmann and R.R. Ernst, *Biochem. Biophys. Res. Commun.*, 78 (1977) 99.
- 265. C. Griesinger, O.W. Sorensen and R.R. Ernst, *J. Magn. Reson.*, 75 (1987) 474.
- 266. L.E. Kay, B. Brooks, S.W. Sparks, D.A. Torchia and A. Bax, *J. Am. Chem. Soc.*, 111 (1989) 5488.
- 267. G.T. Montelione and G. Wagner, *J. Am. Chem. Soc.*, 111 (1989) 5474.
- 268. G.M. Clore, A. Bax and A.M. Gronenborn, *J. Biomol. NMR*, 1 (1991) 13.
- 269. G. Wagner, P. Schmieder, V. Thanabal and G. Wagner, *J. Magn. Reson.*, 93 (1991) 436.
- 270. S.D. Emerson and G.T. Montelione, *J. Magn. Reson.*, 99 (1992) 413.
- 271. C. Griesinger and U. Eggenberger, *J. Magn. Reson.*, 97 (1992) 426.
- 272. G.T. Montelione, S.D. Emerson and Lyons, *Biopolymers*, 32 (1992) 327.
- 273. M. Billeter, D. Neri, G. Otting and Y.Q. Qian, *J. Biomol. NMR*, 2 (1992) 257.
- 274. W. Willker and D. Leibfritz, *J. Magn. Reson.*, 99 (1992) 421.
- 275. J.M. Le Parco, L. McIntyre and R. Freeman, *J. Magn. Reson.*, 97 (1992) 553.
- 276. G.W. Vuister and A. Bax, *J. Am. Chem. Soc.*, 115 (1993) 7772.
- 277. H.B. Olsen, S. Ludvigsen and O.W. Soerensen, *J. Magn. Reson., Ser. A*, A104 (1993) 226.
- 278. K. Bartik and C. Redfield, *J. Biomol. NMR*, 3 (1993) 415.
- 279. F. Fogolari, G. Esposito and P. Viglino, *J. Magn. reson, Ser. A*, A102 (1993) 49.
- 280. H. Kuboniwa, S. Grzesiek, F. Delaglio and Bax, *J. Biomol. NMR*, 4 (1994) 871.
- 281. K.L. Constantine, M.S. Friedrichs and L.M. Mueller, *J. Magn. Reson., Ser. B*, 104 (1994) 62.
- 282. W. Willker and D. Leibfritz, *J. Magn. Reson., Ser. A*, 110 (1994) 95.
- 283. M. Gorlach, M. Wittekind, B.T. Farmer, II, L.E. Kay and L. Mueller, *J. Magn. Reson., Ser. B*, 101 (1993) 194.
- 284. J.M. Schmidt, O.W. Soerensen and R.R. Ernst, *J. Magn. Reson., Ser. A*, 109 (1994) 80.
- 285. G.W. Vuister and A. Bax, *J. Biomol. NMR*, 4 (1994) 193.
- 286. H. Thoegersen and O.W. Soerensen, *J. Magn. Reson., Ser. A*, 110 (1994) 118.

287. M. Tessari, M. Mariani, R. Boelens and R. Kaptein, *J. Magn. Reson., Ser. B*, 108 (1995) 89.
288. G.T. Montelione, M.E. Winkler, P. Rauenbuehler and G. Wagner, *J. Magn. Reson.*, 82 (1989) 198.
289. W. Bermel, K. Wagner and C. Griesinger, *J. Magn. Reson.*, 83 (1989) 223.
290. P. Schmieder, V. Thanabal, L.P. McIntosh, F.W. Dahlquist and G. Wagner, *J. Am. Chem. Soc.*, 113 (1991) 6323.
291. P. Schmieder, M. Kurz and H. Kessler, *J. Biomol. NMR*, 1 (1991) 403.
292. M. Kurz, P. Schmieder and H. Kessler, *Angew. Chem.*, 103 (1991) 1341 (see also *Angew. Chem., Int. Ed. Engl.*, 30 (1991) 1329).
293. A. Bax, D. Max and D. Zax, *J. Am. Chem. Soc.*, 114 (1992) 6923.
294. S. Grzesiek, M. Ikura, G.M. Clore, A.M. Gronenborn and A. Bax, *J. Magn. Reson.*, 96 (1992) 215.
295. G.W. Vuister and A. Bax, *J. Biomol. NMR*, 2 (1992) 401.
296. P.R. Blake, B. Lee, M.F. Summers, M.W.W. Adams, J.B. Park, Z.H. Zhou and A. Bax, *J. Biomol. NMR*, 2 (1992) 527.
297. P. Schmieder and H. Kessler, *Biopolymers*, 32 (1992) 435.
298. U. Wollborn and D. Leibfritz, *J. Magn. Reson.*, 98 (1992) 142.
299. U. Eggenberger, Y. Karimi-Nejad, H. Thuerling, H. Rueterjans and C. Griesinger, *J. Biomol. NMR*, 2 (1992) 583.
300. G.W. Vuister, T. Yamazaki, D.A. Torchia and A. Bax, *J. Biomol. NMR*, 3 (1993) 297.
301. G.W. Vuister, A.C. Wang and A. Bax, *J. Am. Chem. Soc.*, 115 (1993) 5334.
302. G.W. Vuister, F. Delaglio and A. Bax, *J. Biomol. NMR*, 3 (1993) 67.
303. G.W. Vuister and A. Bax, *J. Magn. Reson., Ser. B*, 102 (1993) 228.
304. G. Zhu and A. Bax, *J. Magn. Reson., Ser. A*, 104 (1993) 353.
305. H. Schwalbe, A. Rexroth, U. Eggenberger, T. Geppert and C. Griesinger, *J. Am. Chem. Soc.*, 115 (1993) 7878.
306. H. Schwalbe, W. Samstag, J.W. Engels, W. Bermel and C. Griesinger, *J. Biomol. NMR*, 3 (1993) 479.
307. U. Wollborn, W. Willker and D. Leibfritz, *J. Magn. Reson., Ser. A*, 103 (1993) 86.
308. G.W. Vuister and A. Bax, *J. Am. Chem. Soc.*, 115 (1993) 7772.
309. S. Grzesiek, G.W. Vuister and A. Bax, *J. Biomol. NMR*, 3 (1993) 487.
310. W. Willker, U. Wollborn and D. Leibfritz, *J. Magn. Reson., Ser. B*, 101 (1993) 83.
311. J.C. Madsen, O.W. Soerensen, P. Soerensen and F.M. Poulsen, *J. Biomol. NMR*, 3 (1993) 239.
312. F. del Rio Portilla and R. Freeman, *J. Chem. Soc., Faraday Trans.*, 89 (1993) 4275.
313. E. Kupce and R. Freeman, *J. Magn. Reson., Ser. A*, 104 (1993) 234.
314. S. Seip, J. Balbach and H. Kessler, *J. Magn. Reson., Ser. B*, 104 (1994) 172.
315. M.D. Soerensen, J.J. Led and O.W. Soerensen, *J. Biomol. NMR*, 4 (1994) 135.

- 316. A. Bax, G.W. Vuister, S. Grzesiek, F. Delaglio, A.C. Wang, R. Tschudin and G. Zhu, *Methods Enzymol.*, 239 (1994) 79.
- 317. C. Biamonti, C.B. Rios and B.A. Lyons, *Adv. Biophys. Chem.*, 4 (1994) 51.
- 318. H. Schwalbe, J.P. Marino, G.C. King, R. Wechselberger, W. Bermel and C. Griesinger, *J. Biomol. NMR*, 4 (1994) 631.
- 319. R. Weisemann, H. Rueterjans, H. Schwalbe, J. Schleucher, W. Bermel and C. Griesinger, *J. Biomol. NMR*, 4 (1994) 231.
- 320. G. Zhu, A. Renwick and A. Bax, *J. Magn. Reson., Ser. A*, 110 (1994) 257.
- 321. Y. Karimi-Nejad, J.M. Schmidt, H. Rueterjans, H. Schwalbe and C. Griesinger, *Biochemistry*, 33 (1994) 5481.
- 322. J. Stonehouse and J. Keeler, *J. Mag. Reson.*, A112 (1995) 43.
- 323. C. Griesinger, O.W. Sørensen and R.R. Ernst, *J. Chem. Phys.*, 85 (1986) 6837.
- 324. H. Kuboniwa, S. Grzesiek, F. Delaglio and A. Bax, *J. Biomol. NMR*, 4 (1994) 871.
- 325. M. Gorlach, M. Wittekind and R.A. Beckman, *EMBO J.*, 11 (1992) 3289.
- 326. G.S. Harbison, *J. Am. Chem. Soc.*, 115 (1993) 3026.
- 327. L. Zhu, B.R. Reid and G.P. Drobny, *J. Magn. Reson., Ser. A*, A115 (1995) 206.
- 328. T.J. Norwood, *J. Magn. Reson., Ser. A*, A114 (1995) 92.
- 329. L.E. Kay, D.A. Torchia and A. Bax, *Biochemistry*, 28 (1989) 8972.
- 330. G.M. Clore, P.C. Driscoll, P.T. Wingfield and A.M. Gronenborn, *Biochemistry*, 29 (1990) 7387.
- 331. G. Lipari and A. Szabo, *J. Am. Chem. Soc.*, 104 (1982) 4546.
- 332. G. Lipari and A. Szabo, *J. Am. Chem. Soc.*, 104 (1982) 4559.
- 333. L.E. Kay, L.K. Nicholson, F. Delaglio, A. Bax and D.A. Torchia, *J. Magn. Reson.*, 97 (1992) 359.
- 334. N.A. Farrow, R. Muhandiram, A.U. Singer, S.M. Pascal, C.M. Kay, G. Gish, S.E. Shoelson, T. Pawson, J.D. Forman-Kay and L.E. Kay, *Biochemistry*, 33 (1994) 5984.
- 335. N.A. Farrow, O. Zhang, J.D. Forman-Kay and L. Kay, *J. Biomol. NMR*, 4 (1994) 727.
- 336. A.G. Palmer, III, N.J. Skelton, W.J. Chazin, P.E. Wright and M. Rance, *Mol. Phys.*, 75 (1992) 699.
- 337. J.W. Peng and G. Wagner, *Biochemistry*, 31 (1992) 8571.
- 338. A. Ross, M. Czisch, T. Zink and T.A. Holak, *J. Magn. Reson., Ser. B*, 102 (1993) 314.
- 339. N.J. Skelton, A.G. Palmer III, M. Akke, J. Kordel, M. Rance and W.J. Chazin, *J. Magn. Reson., Ser. B*, 102 (1993) 253.
- 340. A.M. Mandel and A.G. Palmer III, *J. Magn. Reson., Ser. A*, 110 (1994) 62.
- 341. K.T. Dayie and G. Wagner, *J. Magn. Reson., Ser. A*, 111 (1994) 121.
- 342. J.R. Tolman and J.H. Prestegard, *J. Magn. Reson., Ser. B*, 106 (1995) 97.
- 343. J. Boyd, *J. Magn. Reson., Ser. B*, 107 (1995) 279.

344. G.M. Clore, P.C. Driscoll, P.T. Wingfield and A.M. Gronenborn, *Biochemistry*, 29 (1990) 7387.
345. G.M. Clore, P.C. Driscoll, P.T. Wingfield and A.M. Gronenborn, *Biochemistry*, 30 (1991) 312.
346. M.J. Stone, W.J. Fairbrother, A.G. Palmer, III, J. Reizer, M.H. Saier, Jr. and P.E. Wright, *Biochemistry*, 31 (1992) 4394.
347. D.M. Schneider, M.J. Dellwo and A.J. Wand, *Biochemistry*, 31 (1992) 3645.
348. G. Barbato, M. Ikura, L.E. Kay, R.W. Pastor and A. Bax, *Biochemistry*, 31 (1992) 5269.
349. R. Powers, G.M. Clore, S.J. Stahl, P.T. Wingfield and A. Gronenborn, *Biochemistry*, 31 (1992) 9150.
350. I. Chandrasekhar, G.M. Clore, A. Szabo, A.M. Gronenborn and B.R. Brooks, *J. Mol. Biol.*, 226 (1992) 239.
351. C. Redfield, J. Boyd, L.J. Smith, R.A.G. Smith and C.M. Dobson, *Biochemistry*, 31 (1992) 10431.
352. K.L. Constantine, M.S. Friedrichs, A.J. Bell, T.B. Lavoie, L. Mueller and W.J. Metzler, *Fed. Eur. Biochem. Soc.*, 336 (1993) 457.
353. K.L. Constantine, M.S. Friedrichs, V. Goldfarb, P.D. Jeffrey, S. Sheriff and L. Mueller, *Proteins: Struct. Funct. Genet.*, 15 (1993) 290.
354. M.J. Stone, K. Chandrasekhar, A. Holmgren, P.E. Wright and H.J. Dyson, *Biochemistry*, 32 (1993) 426.
355. B.L. Grasberger, A.M. Gronenborn and G.M. Clore, *J. Mol. Biol.*, 230 (1993) 364.
356. M.A.L. Eriksson, H. Berglund, T. Haerd and L. Nilsson, *Proteins: Struct. Funct. Genet.*, 17 (1993) 375.
357. M. Akke, N.J. Skelton, J. Kordel and Palmer, *Biochemistry*, 32 (1993) 9832.
358. J.-W. Cheng, C.A. Lepre and J.M. Moore, *Biochemistry*, 33 (1994) 4093.
359. R.T. Clubb, J.G. Omichinski, K. Sakaguchi, E. Appella, A.M. Gronenborn and G.M. Clore, *Protein Sci.*, 4 (1995) 855.
360. Z. Zheng, J. Czaplicki and O. Jardetzky, *Biochemistry*, 34 (1995) 5212.
361. M. Buck, J. Boyd, C. Redfield, D.A. MacKenzie, D.J. Jeenes, D.B. Archer and C.M. Dobson, *Biochemistry*, 34 (1995) 4041.
362. J.J. Barchi, Jr., B. Grasberger and A.M. Gronenborn, *Protein Sci.*, 3 (1994) 15.
363. N.A. Farrow, O. Zhang, J.D. Forman-Kay and L. Kay, *Biochemistry*, 34 (1995) 868.
364. A.M. Mandel, M. Akke and A.G. Palmer III, *J. Mol. Biol.*, 246 (1995) 144.
365. J.W. Peng and G. Wagner, *Understanding Chem. React.*, 8 (1994) 373.
366. T. Szyperski, P. Luginbuehl, G. Otting and P. Guentert, *J. Biomol. NMR*, 3 (1993) 151.
367. (a) V.Y. Orekhov, K.V. Pervushin and Arseniev, *Eur. J. Biochem.*, 219 (1994) 887.

- (b) M.D. Kemple, P. Yuan, K.E. Nellet, J.A. Fuchs, N. Silva and F.G. Prendergast, *Biophys. J.*, 66 (1994) 211.
- (c) T. Yamazaki, R. Muhandiram and L.E. Kay, *J. Am. Chem. Soc.*, 116 (1994) 8266.
368. K. Linderstrom-Lang, *Chem. Soc. Spec. Publ.*, 2 (1980) 1.
369. G.M. Clore, A. Bax, P.T. Wingfield and A.M. Gronenborn, 29 (1990) 5671.
370. G. Otting, E. Liepinsh, B.T. Farmer, II and K. Wuethrich, *J. Biomol. NMR*, 1 (1991) 20.
371. G. Otting, E. Liepinsh and K. Wüthrich, *Science*, 254 (1991) 97.
372. J.D. Forman-Kay, A.M. Gronenborn, P.T. Wingfield and G.M. Clore, *J. Mol. Biol.*, 220 (1991) 209.
373. T.A. Holak, R. Wütschek and A. Ross, *J. Magn. Reson.*, 97 (1992) 632.
374. S. Grzesiek and A. Bax, *J. Biomol. NMR*, 3 (1993) 627.
375. R.W. Kriwacki, R.B. Hill, J.M. Flanagan, J.P. Caradonna and J.H. Prestegard, *J. Am. Chem. Soc.*, 115 (1993) 8907.
376. G. Gemmecker, W. Jahnke and H. Kessler, *J. Am. Chem. Soc.*, 115 (1993) 11620.
377. S. Campbell-Burk, P. Dommelle and L. Mueller, *J. Magn. Reson.*, 93 (1991) 171.
378. S. Grzesiek and A. Bax, *J. Am. Chem. Soc.*, 115 (1993) 12593.
379. J. Stonehouse, G.L. Shaw, J. Keller and Laue, *J. Magn. Reson., Ser. A*, 107 (1994) 178.
380. L.E. Kay, G.Y. Xu and T. Yamazaki, *J. Magn. Reson., Ser. A*, 109 (1994) 129.
381. T. Yamazaki, S.M. Pascal, A.U. Singer, J.D. Forman-Kay and L.E. Kay, *J. Am. Chem. Soc.*, 117 (1995) 3556.
382. W. Jahnke and H. Kessler, *Angew. Chem., Int. Ed. Engl.*, 34 (1995) 469.
383. B.T. Farmer and R.A. Venters II, *J. Biomol. NMR*, 7 (1995) 59.
384. A.M. Petros, L. Mueller and K.D. Kopple, *Biochemistry*, 29 (1990) 10041.
385. G. Esposito, A.M. Lesk, H. Molinari, A. Motta, N. Niccolai and A. Pastore, *J. Mol. Biol.*, 224 (1992) 659.
386. D.W. Bearden and L.R. Brown, *Chem. Phys. Lett.*, 163 (1989) 432.
387. E.R.P. Zuiderweg, *J. Magn. Reson.*, 89 (1990) 533.
388. M. Ernst, C. Griesinger, R.R. Ernst and W. Bermel, *Mol. Phys.*, 74 (1991) 219.
389. G. Wagner, *J. Magn. Reson.*, 55 (1983) 151.
390. V.V. Krishnan and M. Rance, *J. Magn. Reson. Ser. A*, (1995) 97.
391. G. Bodenhausen, R. Freeman, R. Niedermeyer and D.L. Turner, *J. Magn. Reson.*, 26 (1977) 133.
392. D.I. Hoult and R.E. Richards, *Proc. Roy. Soc. London, Ser. A*, 344 (1975) 311.
393. G. Zhu, D.A. Torchia and A. Bax, *J. Magn. Reson., Ser. A*, 105 (1993) 219.
394. M.A. McCoy and L. Mueller, *J. Magn. Reson.*, 99 (1992) 18.

- 395. J. Keeler, R.T. Clowes, A.L. Davis and E. Laue, *Methods Enzymol.*, 239 (1994) 145.
- 396. A. Bax, T. Mehlkopf, J. Smidt and R. Freeman, *J. Magn. Reson.*, 41 (1980) 502.
- 397. A. Bax and S.S. Pochapsky, *J. Magn. Reson.*, 99 (1992) 638.
- 398. R. Bowtell, R.M. Bowley and P. Glover, *J. Magn. Reson.*, 88 (1990) 643.
- 399. Q. He, W. Richter, S. Vathyam and W.S. Warren, *J. Chem. Phys.*, 98 (1993) 6779.
- 400. W. Richter, S. Lee, W.S. Warren and Q. He, *Science*, 267 (1995) 654.
- 401. P.C.M. van Zijl, M.O.N. Johnson, S. Mori and R.E.H. Hurd, *J. Magn. Reson., Ser. A*, 113 (1995) 265.
- 402. A.A. Maudsley, A. Wokaun and R.R. Ernst, *Chem. Phys. Lett.*, 55 (1978) 9.
- 403. K. Nagayama, K. Wuthrich and R.R. Ernst, *Biochem. Biophys. Res. Commun.*, 90 (1979) 305.
- 404. C.J.R. Counsell, M.H. Levitt and R.R. Ernst, *J. Magn. Reson.*, 64 (1985) 470.
- 405. G.W. Vuister, R. Boelens, R. Kaptein, R.E. Hurd, B. John and P.C.M. Van Zijl, *J. Am. Chem. Soc.*, 113 (1991) 9688.
- 406. A. Bax and G.A. Morris, *J. Magn. Reson.*, 42 (1981) 501.
- 407. R.E. Hurd and B.K. John, *J. Magn. Reson.*, 92 (1991) 658.
- 408. L.E. Kay, P. Keifer and T. Saarinen, *J. Am. Chem. Soc.*, 114 (1992) 10663.
- 409. A.L. Davis, J. Keeler, E.D. Laue and D. Moskau, *J. Magn. Reson.*, 98 (1992) 207.
- 410. J.R. Tolman, J. Chung and J.H. Prestegard, *J. Magn. Reson.*, 98 (1992) 462.
- 411. J. Boyd, N. Soffe, B. John, D. Plant and Hurd, *J. Magn. Reson.*, 98 (1992) 660.
- 412. A.G. Palmer III, J. Cavanagh, R.A. Byrd and M. Rance, *J. Magn. Reson.*, 96 (1992) 416.
- 413. A.G. Palmer III, J. Cavanagh, P.E. Wright and M. Rance, *J. Magn. Reson.*, 93 (1991) 151.
- 414. D.I. Hoult and R.E. Richards, *J. Magn. Reson.*, 24 (1976) 71.
- 415. N. Soffe, J. Boyd and M. Leonard, *J. Magn. Reson., Ser. A*, 116 (1995) 117.
- 416. T. Fujiwara, T. Anai, N. Kurihara and K. Nagayama, *J. Magn. Reson., Ser. A*, 104 (1993) 103.
- 417. M.R. Bendall, *J. Magn. Reson., Ser. A*, 112 (1995) 126.
- 418. E. Kupce and R. Freeman, *J. Magn. Reson., Ser. A*, 115 (1995) 273.
- 419. G. Wider, V. Doetsch and K. Wüthrich, *J. Magn. Reson., Ser. A*, 108 (1994) 255.
- 420. D.L. Mattiello, W.S. Warren, L. Mueller and B.T. Farmer II, *J. Am. Chem. Soc.*, 118 (1996) 3253.
- 421. S. Grzesiek and A. Bax, *J. Am. Chem. Soc.*, 116 (1994) 10196.

- 422. M.A. Markus, K.T. Dayie, P. Matsudaira and Wagner, *J. Magn. Reson., Ser. B*, 105 (1994) 192.
- 423. T. Yamazaki, W. Lee, M. Revington, D.L. Mattiello, F.W. Dahlquist, C.H. Arrowsmith and L. Kay, *J. Am. Chem. Soc.*, 116 (1994) 6464.
- 424. B.T. Farmer II and R.A. Venters, *J. Am. Chem. Soc.*, 117 (1995) 4187.
- 425. R.A. Venters, W.J. Metzler, L.D. Spicer, L. Mueller and B.T. Farmer II, *J. Am. Chem. Soc.*, 117 (1995) 9592.
- 426. S. Grzesiek, P. Wingfield, S. Stahl, J.D. Kaufman and A. Bax, *J. Am. Chem. Soc.*, 117 (1995) 9594.
- 427. W.A. Anderson, W.W. Brey, K.A. Delin, L.F. Fuks, H.D.W. Hill, M.E. Johansson, V.Y. Kotsubo, M. Krivorchko, R.S. Withers and W.H. Wong, in: 36th Experimental Nuclear Magnetic Resonance Conference, 1995.
- 428. S. Spera and A. Bax, *J. Am. Chem. Soc.*, 113 (1991) 5490.
- 429. D.S. Wishart, B.D. Sykes and F.M. Richards, *Biochemistry*, 31 (1992) 1647.
- 430. D.S. Wishart and B.D. Sykes, *Methods Enzymol.*, 239 (1994) 363.
- 431. D.A. Case, H.J. Dyson and P.E. Wright, *Methods Enzymol.*, 239 (1994) 392.
- 432. J.G. Pearson, J.-F. Wang, J.L. Markley and H.-B. Le, *J. Am. Chem. Soc.*, 117 (1995) 8823.
- 433. H.-B. Le, J.G. Pearson, A.C. de Dios and Oldfield, *J. Am. Chem. Soc.*, 117 (1995) 3800.
- 434. J. Kuszewski, J. Qin, A.M. Gronenborn and G. Clore, *J. Magn. Reson., Ser. B*, 106 (1995) 92.
- 435. J. Kuszewski, A.M. Gronenborn and G.M. Clore, *J. Magn. Reson., Ser. B*, 107 (1995) 293.
- 436. B. Celda, C. Biamonti, M.J. Arnau, R. Tejero and G.T. Montelione, *J. Biomol. NMR*, 5 (1995) 161.

### Chapter 3

## Preparation of $^2\text{H}$ , $^{13}\text{C}$ and $^{15}\text{N}$ Isotopically-enriched Proteins for NMR Spectroscopic Investigations

BRIAN J. STOCKMAN

### ABSTRACT

Developments in the field of biotechnology have made it easy and economical to introduce  $^2\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  isotopic labels into proteins either uniformly or selectively. At the same time, an explosion in the number of NMR experiments that can exploit the isotopic labeling to obtain resonance assignments and to extract structural, dynamic and biological information has occurred. A variety of procedures have been developed to isotopically-enrich proteins, most of which involve the use of an organism that expresses or over-expresses the protein of interest. This chapter presents the general details of each and discusses their advantages and disadvantages. Isotopic enrichment using photosynthetic bacteria or algae, bacteria, yeast, and mammalian expression sections will be considered, as well as several methods of enrichment that do not use protein expression vehicles. A brief overview of NMR experiments that exploit isotopic enrichment is also presented.

### 1. INTRODUCTION

From the earliest NMR investigations of proteins it was realized that resonance overlaps were serious obstacles to the complete extraction of spectral information content [1]. While analysis of a few resolved resonances indicated the utility of protein NMR, the challenge of spectral resolution and assignment remained for the bulk of the resonances. In



the 1960s, Jardetzky presented the first general strategy for spectral assignments based on selective deuteration of amino acids [2]. A few years later,  $^{13}\text{C}$  [3] and  $^{15}\text{N}$  [4] NMR were introduced to facilitate protein NMR studies. However, these events were largely overshadowed by the development of two-dimensional  $^1\text{H}$  NMR techniques applied to proteins in the early 1980s [5].

Recently, developments in the field of biotechnology have made it easy and economical to introduce stable isotopic labels into proteins. At the same time, an explosion in the number of NMR experiments that can exploit the isotopic labeling has occurred. A myriad of three- and four-dimensional heteronuclear NMR experiments are now routinely performed on isotopically-enriched proteins [6]. As a consequence, the first step in many protein NMR investigations is to produce  $^{13}\text{C}$  and/or  $^{15}\text{N}$  enriched protein. Since many of the proteins of interest have already been over-expressed to produce protein for other types of studies (enzymology, biological function, etc.), this often involves simply adapting the growth media to  $^{13}\text{C}$ - and/or  $^{15}\text{N}$ -labeled metabolites. This process, as described below, is now routine.

The combination of isotopic labeling and multidimensional, multinuclear NMR has increased the molecular weight limit of proteins amenable to complete resonance assignment and structure determination to around 30 kDa. This is in contrast to using only  $^1\text{H}$  NMR, for which the largest proteins studied in detail have generally been limited to 10 kDa or less. The wide applicability of isotopic labeling in protein NMR is illustrated in Table 1, which lists the more than 120 proteins that have been uniformly  $^{13}\text{C}$ - and/or  $^{15}\text{N}$ -enriched and studied by NMR spectroscopy. The large increase in the number of doubly  $^{13}\text{C}/^{15}\text{N}$ -enriched proteins during the period from 1992 to 1994 is readily apparent.

TABLE 1

Survey of uniformly  $^{13}\text{C}$ - or  $^{15}\text{N}$ -enriched proteins

Protein <sup>a</sup>	Label	Year	Reference <sup>b</sup>
2-phenyloxazolone Fv fragment	$^{13}\text{C}, ^{15}\text{N}$	1991	[7]
434 repressor	$^{13}\text{C}, ^{15}\text{N}$	1992	[8]
acyl-coenzyme A binding protein	$^{13}\text{C}, ^{15}\text{N}$	1993	[9]
Ada	$^{15}\text{N}$	1993	[10]
	$^{15}\text{N}$	1993	[11]
adenylate kinase	$^{13}\text{C}, ^{15}\text{N}$	1993	[12]

Protein <sup>a</sup>	Label	Year	Reference <sup>b</sup>
$\alpha$ 2 repressor	<sup>15</sup> N	1991	[13]
anti-digoxin antibody VL domain	<sup>13</sup> C, <sup>15</sup> N	1993	[14]
Antp homeodomain	<sup>15</sup> N	1991	[15]
apocytochrome b562	<sup>15</sup> N	1991	[16]
Arc repressor	<sup>13</sup> C, <sup>15</sup> N	1993	[17]
azurin ( <i>P. aeruginosa</i> )	<sup>15</sup> N	1992	[18]
azurin ( <i>A. denitrificans</i> )	<sup>15</sup> N	1994	[19]
bacterioopsin	<sup>15</sup> N	1994	[20]
bacteriorhodopsin	<sup>15</sup> N	1992	[21]
barnase	<sup>15</sup> N	1993	[22]
barstar	<sup>15</sup> N	1993	[23]
c-Abl Src homology 2 domain	<sup>15</sup> N	1992	[24]
c-Jun	<sup>15</sup> N	1994	[25]
c-Myb oncoprotein	<sup>13</sup> C, <sup>15</sup> N	1993	[26]
C5a	<sup>15</sup> N	1989	[27]
calbindin D9k	<sup>15</sup> N	1992	[28]
calcineurin B	<sup>13</sup> C, <sup>15</sup> N	1993	[29]
calmodulin	<sup>13</sup> C, <sup>15</sup> N	1990	[30]
	<sup>15</sup> N	1992	[31]
carbonic anhydrase II	<sup>13</sup> C	1991	[32]
CD2 adhesion domain	<sup>15</sup> N	1993	[33]
chymotrypsin inhibitor	<sup>13</sup> C, <sup>15</sup> N	1993	[34]
CMP-KDO synthetase	<sup>15</sup> N	1989	[35]
coat protein(fd)	<sup>15</sup> N	1987	[36]
coat protein (M13)	<sup>15</sup> N	1992	[37]
	<sup>13</sup> C, <sup>15</sup> N	1993	[38]
coat protein (Pf1)	<sup>15</sup> N	1987	[39]
colicin E3 immunity protein	<sup>15</sup> N	1992	[40]
cyclophilin (human)	<sup>13</sup> C, <sup>15</sup> N	1991	[41]
	<sup>15</sup> N	1991	[42]
cyclophilin ( <i>E. coli</i> )	<sup>13</sup> C, <sup>15</sup> N	1993	[43]
cytochrome c2 ( <i>R. rubrum</i> )	<sup>15</sup> N	1987	[44]
cytochrome c2 ( <i>R. capsulatus</i> )	<sup>15</sup> N	1991	[45]
cytochrome c'	<sup>15</sup> N	1987	[44]
cytochrome c553	<sup>13</sup> C, <sup>15</sup> N	1989	[46]
dihydrofolate reductase ( <i>L. casei</i> )	<sup>13</sup> C, <sup>15</sup> N	1991	[47]
dihydrofolate reductase (human)	<sup>13</sup> C, <sup>15</sup> N	1992	[48]
DnaJ	<sup>15</sup> N	1993	[49]
eglin c	<sup>15</sup> N	1992	[50]
elongation factor Tu	<sup>15</sup> N	1991	[51]
elongation factor TFIIS	<sup>15</sup> N	1993	[52]
enzyme IIB cellobiose	<sup>13</sup> C, <sup>15</sup> N	1994	[53]

Protein <sup>a</sup>	Label	Year	Reference <sup>b</sup>
enzyme II manitol,A domain	<sup>13</sup> C, <sup>15</sup> N	1993	[54]
epidermal growth factor	<sup>15</sup> N	1992	[55]
F1F0 ATP synthase c subunit	<sup>15</sup> N	1992	[56]
fatty acid binding protein	<sup>13</sup> C	1993	[57]
ferredoxin (vegetative)	<sup>13</sup> C	1982	[58]
	<sup>13</sup> C, <sup>15</sup> N	1988	[59]
ferredoxin (heterocyst)	<sup>13</sup> C, <sup>15</sup> N	1994	[60]
ferrocytochrome b5	<sup>15</sup> N	1992	[61]
ferrocytochrome c2	<sup>15</sup> N	1990	[62]
fibronectin type III module	<sup>15</sup> N	1992	[63]
FKBP	<sup>15</sup> N	19	[64]
	<sup>13</sup> C, <sup>15</sup> N	1993	[65]
flavodoxin ( <i>A. nidulans</i> )	<sup>15</sup> N	1991	[66]
flavodoxin ( <i>A. 7120</i> )	<sup>13</sup> C, <sup>15</sup> N	1990	[67]
flavodoxin ( <i>D. vulgaris</i> )	<sup>15</sup> N	1993	[68]
	<sup>15</sup> N	1993	[69]
GAL	<sup>15</sup> N	1992	[70]
	<sup>13</sup> C, <sup>15</sup> N	1993	[71]
GATA-1	<sup>13</sup> C, <sup>15</sup> N	1993	[72]
gene V protein	<sup>13</sup> C, <sup>15</sup> N	1994	[73]
glucocorticoid receptor	<sup>15</sup> N	1992	[74]
glucose permease II,A domain	<sup>13</sup> C, <sup>15</sup> N	1992	[75]
glutamine binding protein	<sup>15</sup> N	1992	[76]
glutaredoxin	<sup>15</sup> N	1991	[77]
	<sup>13</sup> C, <sup>15</sup> N	1993	[78]
granulocyte stimulating factor	<sup>13</sup> C, <sup>15</sup> N	1992	[79]
heat shock factor	<sup>13</sup> C, <sup>15</sup> N	1994	[80]
histone H1	<sup>15</sup> N	1993	[81]
hMIP-1β	<sup>13</sup> C, <sup>15</sup> N	1994	[82]
hnRNP A1 protein	<sup>13</sup> C, <sup>15</sup> N	1994	[83]
hnRNP C protein	<sup>13</sup> C, <sup>15</sup> N	1992	[84]
Hpr	<sup>13</sup> C, <sup>15</sup> N	1992	[85]
HU protein	<sup>13</sup> C, <sup>15</sup> N	1994	[86]
IICBGlc	<sup>13</sup> C, <sup>15</sup> N	1994	[87]
IIIGlc	<sup>13</sup> C, <sup>15</sup> N	1991	[88]
interferon γ	<sup>13</sup> C, <sup>15</sup> N	1992	[89]
interleukin-1 receptor antagonist	<sup>13</sup> C, <sup>15</sup> N	1992	[90]
interleukin-1β	<sup>13</sup> C, <sup>15</sup> N	1989	[91]
interleukin-2	<sup>15</sup> N	1992	[92]
interleukin-4	<sup>15</sup> N	1991	[93]
	<sup>13</sup> C, <sup>15</sup> N	1992	[94]
lac repressor headpiece	<sup>15</sup> N	1992	[95]

Protein <sup>a</sup>	Label	Year	Reference <sup>b</sup>
lambda cro repressor	<sup>15</sup> N	1987	[96]
	<sup>15</sup> N	1991	[97]
luciferase	<sup>15</sup> N	1993	[98]
lysozyme (T4)	<sup>15</sup> N	1990	[99]
lysozyme (human)	<sup>15</sup> N	1991	[100]
McPC603 Fv fragment	<sup>13</sup> C, <sup>15</sup> N	1994	[101]
Mnt repressor	<sup>15</sup> N	1991	[102]
MutT	<sup>13</sup> C, <sup>15</sup> N	1993	[103]
ner protein	<sup>15</sup> N	1988	[104]
nuclease	<sup>15</sup> N	1989	[105]
	<sup>13</sup> C, <sup>15</sup> N	1990	[106]
Oct-1 POU DNA binding domain	<sup>15</sup> N	1993	[107]
Oct-3 POU homeodomain	<sup>15</sup> N	1993	[108]
ovomucoid third domain	<sup>15</sup> N	1993	[109]
P22 c2 repressor	<sup>13</sup> C, <sup>15</sup> N	1989	[110]
pancreatic trypsin inhibitor	<sup>15</sup> N	1993	[111]
phospholipase A2	<sup>15</sup> N	1992	[112]
phthalate dioxygenase	<sup>15</sup> N	1989	[113]
PI3K SH3 domain	<sup>15</sup> N	1993	[114]
plastocyanin	<sup>13</sup> C	1983	[115]
profilin (human)	<sup>13</sup> C, <sup>15</sup> N	1993	[116]
profilin I ( <i>Acanthamoeba</i> )	<sup>13</sup> C, <sup>15</sup> N	1993	[117]
protein S	<sup>13</sup> C, <sup>15</sup> N	1994	[118]
ras p21	<sup>13</sup> C, <sup>15</sup> N	1992	[119]
	<sup>15</sup> N	1993	[120]
retinal binding protein	<sup>13</sup> C	1993	[57]
retinoic acid receptor b	<sup>15</sup> N	1992	[121]
ribonuclease T1	<sup>13</sup> C, <sup>15</sup> N	1993	[122]
ribosomal protein S17	<sup>15</sup> N	1993	[123]
RNase H (human HIV)	<sup>13</sup> C, <sup>15</sup> N	1991	[124]
RNase H ( <i>E. coli</i> )	<sup>13</sup> C, <sup>15</sup> N	1991	[125]
rubisco	<sup>15</sup> N	1987	[44]
serine protease PB92	<sup>13</sup> C, <sup>15</sup> N	1994	[126]
Src SH3 domain	<sup>15</sup> N	1993	[127]
Staphylococcal protein A,B domain	<sup>15</sup> N	1990	[128]
Staphylococcal protein A,Z domain	<sup>13</sup> C, <sup>15</sup> N	1993	[129]
sterol carrier protein 2	<sup>15</sup> N	1993	[130]
Streptococcal protein G	<sup>15</sup> N	1992	[131]
stromelysin	<sup>13</sup> C, <sup>15</sup> N	1993	[132]
	<sup>13</sup> C, <sup>15</sup> N	1993	[133]
subtilisin 309	<sup>13</sup> C, <sup>15</sup> N	1994	[134]
TGF- $\alpha$	<sup>15</sup> N	1993	[135]

Protein <sup>a</sup>	Label	Year	Reference <sup>b</sup>
TGF- $\beta$	$^{15}\text{N}$	1993	[136]
thioredoxin (human)	$^{15}\text{N}$	1990	[137]
thioredoxin ( <i>E. coli</i> )	$^{15}\text{N}$	1985	[138]
	$^{15}\text{N}$	1991	[139]
troponin C	$^{15}\text{N}$	1992	[140]
trp repressor	$^{15}\text{N}$	1992	[141]
ubiquitin	$^{15}\text{N}$	1992	[142]
urokinase	$^{13}\text{C}, ^{15}\text{N}$	1992	[143]
villin 14T	$^{13}\text{C}, ^{15}\text{N}$	1994	[144]

<sup>a</sup>For some proteins listed, only a smaller fragment or domain has been labeled.

<sup>b</sup>References listed may not be the earliest occurrence for each protein.

A variety of procedures have been developed to isotopically enrich proteins, most of which involve the use of an organism that expresses or over-expresses the protein of interest. The scope of this chapter is to present the *general* details of each and to discuss their advantages and disadvantages. The specific details of each method, and their adaptation to individual proteins, can be found in the references herein. The first section discusses isotopic enrichment using photosynthetic bacteria or algae, bacteria, yeast, and mammalian expression systems. This order roughly parallels development of these techniques for protein enrichment. Enrichment methods that do not use protein expression vehicles are also presented. The next two sections present a brief description of methods used to determine the level of isotopic enrichment and a short survey, with key references, to some of the NMR experiments that exploit isotopic enrichment. The chapter concludes with an outlook for future developments.

## 2. ISOTOPIC ENRICHMENT METHODOLOGIES

The two most important parameters defining protein isotopic enrichment are the *level of enrichment* and the *pattern of enrichment*. The level of enrichment refers to the percentage of the desired isotope incorporated into a given position in the protein. Pattern of enrichment refers to whether the labeling is residue-type or residue-site specific, or is uniform throughout the protein.

A few NMR experiments make use of enrichment levels less than 100%. The most common are those designed to obtain stereospecific

assignments and are described in Section 2.2.2. For most NMR experiments, however, sensitivity is maximized by 100% isotopic enrichment. Isotope-editing experiments on selectively enriched proteins rely, by definition, on the presence of  $^{13}\text{C}$  or  $^{15}\text{N}$  labels, so the sensitivity increases linearly with the level of enrichment. Given sufficient quantity of protein, however, lower levels of enrichment can be used in isotope-editing experiments. This is not the case, however, for triple-resonance experiments that contain numerous magnetization transfer steps involving  $^{13}\text{C}$ – $^{13}\text{C}$  and  $^{13}\text{C}$ – $^{15}\text{N}$  one-bond J-couplings. Here the effect of lower levels of enrichment is magnified, since the occurrence of a  $^{12}\text{C}$  or  $^{14}\text{N}$  at *any* magnetization transfer step results in loss of sensitivity. For instance, an experiment on a uniformly 90%  $^{13}\text{C}/^{15}\text{N}$ -enrichment protein that contains three such magnetization transfer steps will be about 25% less sensitive than on the same protein that is 100%  $^{13}\text{C}/^{15}\text{N}$ -enriched.

The pattern of enrichment can be either uniform or selective, depending on the information desired. For many proteins, it has become standard to record a battery of NMR experiments on uniformly  $^{15}\text{N}$ - or  $^{13}\text{C}/^{15}\text{N}$ -enriched samples. These data sets are then supplemented with experiments recorded on samples enriched only at certain residue types in order to facilitate the sequential assignment process. While uniform labeling can potentially lead to a complete structural investigation, it has the disadvantage that particular regions of the protein cannot be focused upon without assigning all of the resonances first. In contrast, residue-type labeling permits attention to be focused on selected areas of the protein, such as the active site or the receptor recognition site, without interference from resonances arising from other parts of the protein. Recently, site-specific labeling has become possible using several enrichment vehicles, allowing even greater spectral simplification.

For the discussion here, isotopic enrichment methodologies will be categorized in terms of the different organisms that can be used. In each case, various levels and patterns of isotopic enrichment are possible, depending on the NMR experiments to be performed and the information desired. The rationale and protocols used to prepare various levels and patterns of isotopic enrichment will be discussed in each case.

### ***2.1. Photosynthetic bacteria***

Proteins produced in large quantities by photosynthetic bacteria or algae have been isotopically enriched with  $^2\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$ . Because the organisms can grow on media consisting of small molecule precursors

such as  $^{13}\text{CO}_2$  and  $\text{K}^{15}\text{NO}_3$ , this method of enrichment is very economical. Furthermore, the unused biomass can be used as an inexpensive source of labeled amino acid hydrolysate for use in other expression systems such as bacteria, yeast or mammalian cells [106,145].

Photosynthetic bacteria were first used to prepare uniformly  $^2\text{H}$ -labeled proteins containing several types of amino acids selectively protonated [146]. This was accomplished by providing the selected amino acids in their protonated form in an otherwise deuterated amino acid media. Enrichment levels greater than 80% were reported. The resulting NMR spectra were drastically simplified, and allowed the systems of interest to be analyzed in greater detail. Preparation of fully deuterated algae has also been described [147].

Uniformly  $^{13}\text{C}$ - and  $^{15}\text{N}$ -labeled proteins have also been obtained from photosynthetic bacteria. Ferredoxin [58] and plastocyanin [115], obtained from the cyanobacterium *Anabaena variabilis*, were the first two examples of enrichment with  $^{13}\text{C}$ . More recently, a variety of proteins from *Anabaena* 7120 [46,59] and *Rhodospirillum rubrum* [44] have been prepared in their  $^{15}\text{N}$ - and/or  $^{13}\text{C}/^{15}\text{N}$ -labeled forms in this fashion.

Except for the selective deuteration discussed above, photosynthetic bacteria have not been exploited to produce selectively enriched proteins. This may stem from the ease and economy with which these organisms can be uniformly enriched, especially considering the cost and availability of  $^{13}\text{C}$ - and  $^{15}\text{N}$ -labeled amino acids at the time of these studies. In addition, proteins enriched using this method have been limited to those produced naturally and in high quantities by the photosynthetic bacteria, such as photosynthetic electron-transport proteins. Thus this method of enrichment is not generally applicable to other proteins. Given the overriding ease and utility of bacterial expression, this is not likely to change. Indeed, several of the proteins discussed above have now been cloned and over-expressed in bacteria, making their production in algae obsolete.

## 2.2. Bacteria

Bacterial expression is by far the most common means of preparing selectively and uniformly isotopically-enriched proteins (see references in Table 1). Originally, bacterial proteins produced during the normal metabolism of the organism were enriched. In pioneering experiments, Markley and coworkers [148] and Browne and coworkers [3] demonstrated the selective incorporation of  $^2\text{H}$  into *Staphylococcal* nuclease

and  $^{13}\text{C}$  into *E. coli* tryptophan synthetase a subunit, respectively. With the advent of molecular biology techniques, the use of bacterial expression systems to over-express isotopically-enriched non-endogenous proteins has exploded. In this section, a survey of the techniques used to enrich proteins over-expressed in bacteria will be given.

The most common bacteria used for protein enrichment is *E. coli* (see references in Table 1). Use of this organism has several key advantages over other bacteria that have also been used. Advantages include: the ease and widespread use of *E. coli* coupled with plasmid or bacteriophage expression vectors to obtain unenriched protein for other purposes, such as enzymology studies or biological screening assays (i.e. the protein has already been expressed in *E. coli* by the time the NMR spectroscopist comes along); *E. coli* can efficiently take up and incorporate into proteins exogenously supplied isotopically-labeled compounds; the amino acid metabolic pathways of *E. coli* are well understood, allowing them to be genetically blocked or altered if required to obtain the desired enrichment pattern; the yield of protein per liter of medium is often in the tens of milligrams, allowing cost-efficient production of the amounts of material necessary for NMR studies [149]. Proteins expressed in *E. coli* can be enriched either uniformly or selectively, depending on the information desired. Expression of  $^2\text{H}$  proteins in bacteria has been recently reviewed [150]. Only  $^{13}\text{C}$  and  $^{15}\text{N}$  enrichment will be discussed here.

Although bacterial expression is widely applicable to many proteins, several eucaryotic proteins have been found to not fold properly in bacteria (lysozyme for example [100]), and therefore cannot be enriched using this technique. Lack of glycosylation capabilities by the host *E. coli* may be responsible for the inability of some eucaryotic proteins to express well. On the other hand, production of non-glycosylated proteins is advantageous since heterogeneous glycosylation may impair the interpretation of the NMR data. Provided that no functional differences exist between the physiologically glycosylated version of a protein and the non-glycosylated bacterially-derived version, lack of glycosylation is a distinct advantage of bacterial expression.

### 2.2.1. Uniform enrichment

Media to produce uniformly isotopically-enriched proteins in *E. coli* can be divided into two categories: minimal media [151], containing the regular carbon and/or nitrogen sources substituted with their  $^{13}\text{C}$  and/or  $^{15}\text{N}$  analogues, and rich media, containing the  $^{13}\text{C}$  and/or  $^{15}\text{N}$  analogues



of the amino acids present. In the case of minimal media, the labeled precursors are  $^{15}\text{NH}_4\text{Cl}$  at a level of 1 gram per liter and  $[^{13}\text{C}_6]\text{glucose}$  at a level of 2–3 grams per liter. Both compounds are readily available commercially. For rich media, the amino acids can be supplied as a hydrolysate of bacteria- or algae-derived proteins or as a mixture prepared from individual amino acids. The former method is the most cost-effective, as the hydrolysate can be prepared in-house or can be purchased commercially. The latter is more costly, and is not yet feasible for doubly  $^{13}\text{C}/^{15}\text{N}$ -enriched proteins since the complete set of  $^{13}\text{C}/^{15}\text{N}$ -enriched amino acids is not available. The latter method, however, has the advantage, discussed below, of allowing the incorporation of selected labeled amino acid types into the protein.

The level of enrichment obtained correlates with the level of enrichment contained in the labeled small molecules supplied in the growth media. Typically, enrichment levels greater than 98% are obtained for both  $^{13}\text{C}$  and  $^{15}\text{N}$  when fully-enriched precursors are used. The only dilution of total label that occurs is attributable to breakdown products of the initial unlabeled bacteria used to seed the isotopically-enriched growth media. Because the bacteria must synthesize all amino acids from scratch,  $^{13}\text{C}$  and  $^{15}\text{N}$  are both incorporated uniformly, with backbone and side chain atoms being labeled for each amino acid. In addition, any cofactors or ligands produced by the *E. coli* will be enriched, provided that they have not been supplied in their unlabeled form in the growth media. No functional differences between unenriched and  $^{13}\text{C}$ - or  $^{15}\text{N}$ -enriched protein have been reported.

### 2.2.2. Selective enrichment

For both large and small proteins, it is often of interest to enrich only certain parts of the protein, thereby simplifying the resulting NMR spectra and its interpretation. In the case of  $^{15}\text{N}$ , it is often advantageous to incorporate one or a subset of amino acids in their labeled form, allowing residue-type assignment of these resonances. This often provides starting points for sequential assignments of all resonances [152, 153]. For  $^{13}\text{C}$ , it is often useful to selectively label side chains in a defined manner in order to make stereospecific assignments [110], or to combine selective  $^{13}\text{C}$  and  $^{15}\text{N}$  labels in order to make sequential resonance assignments [154].

Selective labeling with  $^{15}\text{N}$  in *E. coli* has been exquisitely detailed recently by McIntosh and Dahlquist [149]. Only general considerations will be discussed here. Selective labeling with  $^{15}\text{N}$  in the amide backbone

position is accomplished by supplying the desired amino acids in their  $^{15}\text{N}$ -enriched form in the mixture of amino acids used in the growth media. Typical amounts of each amino acid used in the media are discussed by McIntosh and Dahlquist [149]. The caveat to this approach is that one has to consider not only what happens to the  $^{15}\text{N}$  label when the  $^{15}\text{N}$ -enriched amino acid enters into the metabolism of the host *E. coli*, but also the possibility that unenriched versions of the supplied  $^{15}\text{N}$ -labeled amino acids will be created by the host.

When a given amino acid is introduced in the  $^{15}\text{N}$ -enriched form, the label can be scrambled by the host's metabolism through a variety of transamination reactions. This could potentially lead to a reduced level of label in the desired residue type, as well as finite amounts of label appearing in different types of residues. *E. coli* amino acid metabolism, as well as a list of *E. coli* transaminases, can be found in McIntosh and Dahlquist [149]. An understanding of all the potential sources and fates of an introduced  $^{15}\text{N}$ -labeled amino acid is required to interpret the resulting enrichment pattern. Scrambling of label can be substantially reduced by a combination of two procedures. First, sufficient quantities of all non-enriched amino acids should be present in the growth media, thereby suppressing the *E. coli* from biosynthesizing these amino acids from the labeled amino acid. Second, transaminase-deficient or amino acid auxotrophic strains of *E. coli* should be used to suppress the loss of label by these mechanisms.

McIntosh and Dahlquist [149] also point out that knowledge of *E. coli* metabolism can be used in several advantageous ways. More than one type of residue can be labeled by supplying a single  $^{15}\text{N}$ -enriched amino acid if the enriched amino acid is the metabolic precursor for several other amino acids, and the other amino acids are not supplied in the growth media. This may provide a cost-effective labeling strategy, as well as reduce the number of different residue-type labeled samples necessary to extract the desired information. Likewise, patterns of  $^{15}\text{N}$ -label dilution may be recognizable, as for example in human dihydrofolate reductase [48], and thus may provide more information than if the label had not been diluted.

Selective  $^{15}\text{N}$  enrichment of amino acid side chains is also feasible. The histidine side-chain nitrogen atoms in a-lytic protease have been labeled using a strain of myxobacter auxotrophic for histidine [155]. Selective enrichment of the side-chain tryptophan nitrogens, using an auxotrophic strain of bacteria, and the side-chain arginine nitrogens, using biosynthetic pathways, has been reported for T4 lysozyme [149].

Selective enrichment of proteins with  $^{13}\text{C}$ , as with  $^{15}\text{N}$ , was first done to simplify spectral analysis and assignment, and to provide probes of biological function. For instance, pioneering work by Browne et al. [3] demonstrated that histidine residues could be selectively enriched in tryptophan synthetase a subunit. Many examples of using this technique to focus the NMR studies on specific sites in a protein have appeared. Perhaps the best studied protein from a selective  $^{13}\text{C}$ -enrichment standpoint is *Streptomyces* subtilisin inhibitor, which has been prepared with a variety of specific labels (for example see Ref. [156]). Selective  $^{13}\text{C}$  enrichment in combination with selective  $^{15}\text{N}$  enrichment has also been developed as a strategy to make sequential assignments in proteins based on unique dipeptides [153,154].

Selective  $^{13}\text{C}$  enrichment necessary for the above experiments is done in an analogous fashion to  $^{15}\text{N}$  enrichment in *E. coli*. The desired  $^{13}\text{C}$  amino acid is substituted in the media for the unlabeled version. In contrast to selective  $^{15}\text{N}$  enrichment, dilution of label through transamination reactions is not a major problem. As a general rule,  $^{13}\text{C}$  labels scramble to a much lesser degree than  $^{15}\text{N}$  labels [136].

The potential to use bacterial expression to incorporate *residue-site* specific, in contrast to *residue-type* specific, isotope labels into proteins has recently been demonstrated [157]. Utilizing the general technique [158] of in vitro suppression of a TAG nonsense mutation, [ $^{13}\text{C}$ ]alanine was introduced in position 82 of T4 lysozyme. Because the nonsense mutation is present only at position 82, this is the only alanine that is enriched in the protein. By contrast, residue-type labeling would have resulted in labeling of all alanine positions indiscriminately. Residue-site specific labeling provides a method to put an isotopic label at any position in a protein, and should be widely applicable to many systems, especially the active sites or ligand binding sites of larger proteins.

A different pattern of selective  $^{13}\text{C}$  enrichment has been developed with the aim of assisting in the stereospecific assignment of side-chain methyl groups and to support assignment of  $^1\text{H}$  spin systems. Termed nonrandom or biosynthetically directed fractional  $^{13}\text{C}$ -labeling [110, 159], proteins are obtained from *E. coli* that has been grown on a mixture of 10% [ $^{13}\text{C}_6$ ]glucose and 90% unlabeled glucose as the sole carbon source. Assuming normal *E. coli* metabolism of the glucose into amino acids and disregarding the natural  $^{13}\text{C}$  abundance of 1.1%, the probability of adjacent  $^{13}\text{C}$  atoms in the protein is 1% if the two atoms involved originate from *different* glucose molecules. If the adjacent carbon atoms are derived from the *same* original glucose molecule, the

probability is 10% that they will both be  $^{13}\text{C}$ -labeled. For valine and leucine side-chain methyl groups, both situations exist. The pro-R methyl group and the adjacent methine carbon are derived from the same glucose-derived pyruvate molecule, while the pro-S methyl group and the adjacent methine carbon originate from different glucose-derived pyruvate molecules. Observation of the pattern of enrichment allows the pro-R and pro-S methyl groups to be distinguished and thus to be stereospecifically assigned. This method has also been shown to be generally applicable in the assignment of all types of methyl groups, and to be supportive of  $^1\text{H}$  spin system identification based on the coupling patterns observed [160].

### 2.3. Yeast

Although bacterial expression is by far the most common method for producing enriched proteins, some proteins cannot be obtained in this fashion. A variety of proteins that do not fold correctly in bacterial expression systems, however, have been successfully expressed in yeast. Techniques for gene over-expression and protein production in yeast are well known. For a good review see *Methods in Enzymology* Volume 194, especially the articles by Rine [161] and Moir and Davidow [162]. Proteins have been obtained from yeast grown on a minimal media similar to that used for bacterial growth, as well as from yeast grown on rich media containing a mixture of amino acids. As with bacteria, growth media for yeast can easily be adapted for  $^{13}\text{C}$  and  $^{15}\text{N}$  enrichment by substituting  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled precursors for the normal small molecule precursors.

In contrast to bacterial expression, yeast expression presents the potential for protein glycosylation since the eucaryotic yeast has the machinery to glycosylate proteins. This is especially true for proteins that are secreted by the yeast, and may occur to a lesser extent for proteins that remain in the cytoplasm of the yeast. Production of proteins using the yeast secretory pathway increases the likelihood of correct folding with disulfide bonds, increases the likelihood of the proper asparagine residues receiving N-linked glycosylation, and removes the desired protein from the bulk of the yeast proteinases [162].

Three proteins have been uniformly enriched using yeast expression systems: [ $^{15}\text{N}$ ]human lysozyme [100], [ $^{15}\text{N}$ ]fibronectin cell adhesion type III module [63], and [ $^{15}\text{N}$  and  $^{13}\text{C}/^{15}\text{N}$ ]human interleukin-4 [94]. In each study, a strain of *Saccharomyces cerevisiae* was used. The growth media

used for the preparation of lysozyme and fibronectin module contained either  $^{15}\text{NH}_4\text{OH}$  or  $(^{15}\text{NH}_4)_2\text{SO}_4$  as the sole nitrogen sources, while the growth media used for the production of interleukin-4 contained  $^{15}\text{N}$  and/or  $^{13}\text{C}$  labeled amino acid hydrolysate supplemented with  $^{15}\text{NH}_4\text{Cl}$  and/or  $[^{13}\text{C}_6]\text{glucose}$ . For human lysozyme, the yield reported was about 5 mg/l culture, with an enrichment level measured from peak intensities near 90%. Both of these values are slightly lower than typically obtained using bacterial expression. For interleukin-4, the reported level of enrichment was 96.6% for both  $^{13}\text{C}$  and  $^{15}\text{N}$ , based on mass spectrometry.

In the case of human interleukin-4, two potential glycosylation sites, asparagine-42 and asparagine-109, were mutated to aspartic acid residues. Based on mass spectrometry, no glycosylation or other posttranslational modification occurred in the product protein. No mention was made of glycosylation or lack thereof in lysozyme or in the fibronectin domain.

## **2.4. Mammalian**

An alternative to yeast for the production of proteins that cannot be expressed using bacterial systems is mammalian cells. Expression in mammalian cells has been used to produce numerous soluble proteins (for a review see *Methods in Enzymology* Volume 185). Hybridoma cells have been used to express antibodies [163] and fragments thereof [164,165], as well as a variety of other proteins. Other types of mammalian cells have also been used to express proteins of interest. Techniques to enrich proteins expressed in hybridoma cells will be discussed first. In the second section, recently-developed general protocols to enrich proteins expressed in Chinese hamster ovary (CHO) cells will be discussed.

### **2.4.1. Hybridoma cells**

Hybridoma cells (AN02) were first utilized to obtain  $^2\text{H}$ -labeled antibody Fab fragment by Anglister et al. [166]. In this investigation, selected amino acids were supplied in their deuterated form in the serum-containing (1%) growth media, with all other amino acids being supplied in the protonated form. The incorporation of deuterated amino acids provided detailed information about the residues in and around the antigen binding site.

Selective  $^2\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  labeling has also been applied by Arata and coworkers to investigate entire antibodies [167] and Fv fragments

[168,169]. Switch-variant cell lines 27–44 (IgG1), 27–13.6 (IgG2a), 27–1B10.7 (IgG2a.s) and 27–35.8 (IgG2b) were utilized. The media used was serum-free, with the desired labeled amino acids being substituted for unlabeled ones. The authors noted that some transaminase-derived isotopic scrambling occurred for all  $^{15}\text{N}$  amino acids used except for lysine and threonine. Transamination could largely be suppressed by the addition of  $\beta$ -chloro-L-alanine to the growth media. Incorporation of selective labels has allowed a detailed study of the antigen binding site structure and dynamics.

Recently, a method to prepare uniformly enriched urokinase expressed in hybridoma Sp2/0 cells [170] has been presented [143]. Like the selective enrichment discussed above, labeled amino acids are used in the growth media. However, in order to make the method cost-effective, the amino acids were derived from algal or bacterial hydrolysates. It was found that the amino acid mixture, without further purification, was unable to support Sp2/0 cell growth, presumably because of endotoxins or other toxins carried over from the bacteria or algae. The authors found that Sp2/0 cell growth could be supported when the toxins were removed by dialysis and affinity chromatography. A growth medium consisting of 5% dialyzed heat-treated serum supplemented with the toxin-less amino acid hydrolysate and two synthesized labeled amino acids lost during the acid hydrolysis (glutamine and cysteine) was found to support Sp2/0 growth to the same level as a similar media reconstituted with commercially available unlabeled amino acids. Based on NMR spectra of urokinase purified from the Sp2/0 cells, enrichment levels achieved were determined to be greater than 95%. The total yield of urokinase was 30 mg/l culture. The estimated cost per liter of media was US \$1,000 for  $^{15}\text{N}$  media and \$4,000 for  $^{13}\text{C}/^{15}\text{N}$  media. Curiously, the authors found that enzymatically hydrolyzed bacterial or algal amino acids could not support Sp2/0 cell growth, but that acid hydrolyzed amino acids could.

#### 2.4.2. CHO cells

Hansen et al. [143] also reported the ability of their Sp2/0 cell media to support the growth of CHO cells, stating that CHO cells grew to the same density in labeled or unlabeled media. Since many proteins have been expressed in CHO cells, this provides a widely-applicable protocol to prepare labeled proteins using mammalian expression systems.

Indeed, Archer et al. [136] have reported the independent development of a protocol to isotopically enrich transforming growth factor  $\beta$ 1

expressed in CHO cells [171]. In this investigation, TGF- $\beta$ 1 was uniformly  $^{15}\text{N}$  enriched and selectively  $^{13}\text{C}$  and  $^{15}\text{N}$  enriched. The media used was 1% dialyzed heat-inactivated serum supplemented with the desired  $^{13}\text{C}$  and/or  $^{15}\text{N}$  metabolites. For uniform  $^{15}\text{N}$  labeling,  $^{15}\text{N}$ -labeled choline and uniformly  $^{15}\text{N}$ -labeled amino acids were used. For selective  $^{13}\text{C}/^{15}\text{N}$  labeling, the appropriate labeled amino acids were substituted in the growth media. The yield of TGF- $\beta$ 1 was approximately 1 g/l culture, with a cost per liter of \$1,000. The level of enrichment obtained was greater than 90%, as judged from the NMR spectra. Uniformly  $^{15}\text{N}$  and partially  $^{13}\text{C}$  double-labeled, as well as selectively [ $^{15}\text{N}$ ]lysine labeled, human CD2 adhesion domain has also been prepared from CHO cells [33].

Archer et al. [136] also analyzed the scrambling or lack thereof of the selective amino acid labels. For  $^{15}\text{N}$ , they found that the levels of selective incorporation correlated well with the levels contained in the growth media, except for glycine and serine, which scrambled between themselves to a high extent. High levels of scrambling were also observed between leucine, valine and isoleucine. Low levels of scrambling were observed for several other amino acids as well. Limited scrambling of label between asparagine and glutamine side chain amides was also observed. By contrast, the authors did not observe scrambling of the  $^{13}\text{C}$  labels.

Supplementing the growth media with each individual  $^{13}\text{C}$ -labeled amino acid in order to prepare uniformly  $^{13}\text{C}$ -enriched TGF- $\beta$ 1 was too costly. Until the availability and price of  $^{13}\text{C}$ -enriched amino acids makes this economically feasible, the use of toxin-free amino acid hydrolysate will be the most cost-effective method of preparing doubly  $^{13}\text{C}/^{15}\text{N}$ -enriched proteins in mammalian expression systems.

## ***2.5. Non-biosynthetic methods of enrichment***

Several techniques that do not require biosynthesis have been used to incorporate  $^2\text{H}$ ,  $^{13}\text{C}$  or  $^{15}\text{N}$  labels into proteins. These include solid-phase synthesis, chemical exchange and enzymatic semisynthesis.

For smaller proteins, solid-phase synthesis techniques have been used to selectively label specific residues [172]. Unlike biosynthetic labeling, where each occurrence of the specific type of amino acid in the sequence is labeled, solid-phase synthesis permits labeling of any desired combination of site-specific residues. Preparation of uniformly-labeled small molecular weight proteins with this method, while possible, is not cost-effective. Larger proteins are not amenable at all to this technique.

Chemical exchange [173] simply involves the exchange of amide or histidine  $^1\text{H}^{\text{e1}}$  protons for deuterons (or *vice versa*). Enzymatic semisynthesis [174] involves the use of a specific enzyme to introduce a label into the desired position of the protein. The former technique is widely applicable to proteins, while the latter is very specific for the system being investigated.

### 3. DETERMINATION OF ENRICHMENT LEVEL

The level of enrichment observed in the final product proteins resembles the isotopic composition of the growth media. For uniform labeling, enrichment is uniformly random. Various methods can be used to determine the level of enrichment. Comparing the intensities of the  $^1J_{\text{XH}}$  ( $\text{X} = ^{13}\text{C}$  or  $^{15}\text{N}$ ) satellite resonances to the residual unsplit resonance of a well-resolved, undecoupled  $^1\text{H}$  signal provides a quick estimate. Mass spectrometry on the enriched protein provides a more accurate answer. For selective labeling, isotopic dilution or scrambling, as discussed above may occur, and may create a variety of different enrichment levels at different residue or atom types. The total amount of enrichment should be obtainable using mass spectrometry. Intensity measurements and resonance assignments could then be used to determine the level of enrichment at specific sites or residue types.

### 4. NMR SPECTROSCOPY OF ISOTOPICALLY ENRICHED PROTEINS

NMR experiments that exploit isotopic labeling fall into two general classes: those designed to obtain resonance assignments, and those designed to extract structural, dynamic and biological information. Selective and uniform isotopic enrichment are applicable to both.

Selective isotope labeling can assist the assignment process by allowing resonances to be classified according to residue type. Assignment of T4 lysozyme is an excellent example of this technique [99]. Selective double labeling, where enrichment of the carbonyl carbon of residue X and the amide nitrogen of residue Y results in sequential assignment when X-Y is a unique dipeptide in the protein [154], has found widespread applicability. Application of isotope-editing, whereby the  $^{13}\text{C}$  or  $^{15}\text{N}$  label is used to edit the  $^1\text{H}$  NMR spectrum [175], allows the experimenter to focus structural or dynamic studies on one part of the protein.

Uniform  $^{15}\text{N}$  enrichment can be used to support an  $^{15}\text{N}$ -directed assignment strategy that relies on concerted analysis of three-dimen-



sional  $^1\text{H}$ - $^{15}\text{N}$  NOESY-HMQC and  $^1\text{H}$ - $^{15}\text{N}$  TOCSY-HMQC data [91]. Many of the proteins in Table 1 that are only labeled with  $^{15}\text{N}$  have had their resonances assigned in this fashion. Uniform  $^{13}\text{C}/^{15}\text{N}$  enrichment makes possible a battery of triple-resonance experiments that use scalar connectivities, not NOEs, to make sequential assignments [30, 129]. Backbone assignments are then extended to side-chain assignments through the use of  $^{13}\text{C}$  TOCSY experiments [176,177]. Beyond resonance assignments,  $^{13}\text{C}$  and  $^{15}\text{N}$  labeling permits three- and four-dimensional NOESY experiments to be recorded, increasing spectral resolution and concomitantly the number of NOEs that can be uniquely identified [178]. Likewise, heteronuclear coupling constants can be obtained, using the  $^{13}\text{C}$  or  $^{15}\text{N}$  label to increase resolution and to create the desired coupling patterns [179]. Finally,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts themselves may be useful structural parameters [180].

## 5. FUTURE PERSPECTIVES

The number of proteins that have been isotopically enriched increases monthly. Given the widespread popularity of bacterial over-expression, this trend will continue. As the demand for  $^{13}\text{C}/^{15}\text{N}$ -enriched amino acids,  $^{13}\text{C}/^{15}\text{N}$ -enriched amino acid hydrolysate,  $^{15}\text{NH}_4\text{Cl}$  and [ $^{13}\text{C}_6$ ]glucose rises, their availability and cost will improve. For example, in 1990 the cost for a gram of [99%  $^{13}\text{C}_6$ ]glucose was approximately \$650. The cost per gram today is below \$400. Reduced cost may be especially relevant for facilitating studies using multiply  $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -enriched proteins [181].

With the introduction of protocols for yeast and mammalian cell expression of proteins, the potential to enrich a still more diverse set of proteins than listed in Table 1 exists. This will be especially important for eucaryotic proteins that do not express well in bacterial systems. This will most likely lead to the assignment and structure determination of glycosylated proteins [33]. Since the carbohydrate moieties will be enriched as well, assignments and structural information may not be limited to only the polypeptide residues.

Finally, the upper limit of molecular weight amenable to in-depth study will continue to rise. Although this is highly dependent on intrinsic properties of the protein in question, isotopic enrichment combined with development of methods to improve the solution properties of proteins [182,183] should extend the applicability of these techniques to proteins in the 30–40 kDa range in the near future. Just as important,

larger proteins (>50 kDa) will continue to be investigated using selective labeling techniques. The recently developed procedure of using suppressor tRNAs to incorporate site-specific labels will provide access to the active sites or surface ligand-binding sites of many interesting proteins.

## REFERENCES

1. M. Saunders, A. Wishnia and J.G. Kirkwood, *J. Am. Chem. Soc.*, 79 (1957) 3289.
2. O. Jardetzky, *Proc. Int. Symp. Nuclear Magn. Res.*, Tokyo, Japan, N-3-14 (1965).
3. D.T. Browne, G.L. Kenyon, E.L. Packer, H. Sternlicht and D.M. Wilson, *J. Am. Chem. Soc.*, 95 (1973) 1316.
4. D. Gust, R.B. Moon and J.D. Roberts, *Proc. Natl. Acad. Sci. USA*, 72 (1975) 4696.
5. K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.
6. G.M. Clore and A.M. Gronenborn, *Science (Washington, D.C.)*, 252 (1991) 1390.
7. L. Riechmann, J. Cavanagh and S. McManus, *FEBS Lett.*, 287 (1991) 185.
8. D. Neri, G. Wider and K. Wüthrich, *FEBS Lett.*, 303 (1992) 129.
9. B.B. Kragelund, K.V. Andersen, J.C. Madsen, J. Knudsen and F.M. Poulsen, *J. Mol. Biol.*, 230 (1993) 1260.
10. H. Sakashita, T. Sakuma, T. Ohkubo, M. Kainosho, K. Sakumi, M. Sekiguchi and K. Morikawa, *FEBS Lett.*, 323 (1993) 252.
11. L.C. Myers, G.L. Verdine and G. Wagner, *Biochemistry*, 32 (1993) 14089.
12. I.-J.L. Byeon, H. Yan, A.S. Edison, E.S. Mooberry, F. Abildgaard, J. L. Markley and M.-D. Tsai, *Biochemistry*, 32 (1993) 12508.
13. C.L. Phillips, A.K. Vershon, A.D. Johnson and F.W. Dahlquist, *Genes Dev.*, 5 (1991) 764.
14. K.L. Constantine, V. Goldfarb, M. Wittekind, M.S. Friedrichs, J. Anthony, S.-C. Ng and L. Mueller, *J. Biomol. NMR*, 3 (1993) 41.
15. G. Otting, Y.Q. Qian, M. Billeter, M. Muller, M. Affolter, W.J. Gehring and K. Wüthrich, *EMBO J.*, 9 (1990) 3085.
16. Y.Q. Feng, A.J. Wand and S.G. Sligar, *Biochemistry*, 30 (1991) 7711.
17. M.J.M. Burgering, R. Boelens, M. Caffrey, J.N. Berg and R. Kaptein, *FEBS Lett.*, 330 (1993) 105.
18. M. van de Kamp, G.W. Canters, S.S. Wijmenga, A. Lommen, C.W. Hilbers, H. Nar, A. Messerschmidt and R. Huber, *Biochemistry*, 31 (1992) 10194.
19. C.W.G. Hoitink, P.C. Driscoll, H.A.O. Hill and G.W. Canters, *Biochemistry*, 33 (1994) 3560.

20. K.V. Pervushin, V.Y. Orekhov, A.I. Popov, L.Y. Musina and A.S. Arseniev, *Eur. J. Biochem.*, 219 (1994) 571.
21. V.Y. Orekhov, G.V. Abdulaeva, L.Y. Musina and A.S. Arseniev, *Eur. J. Biochem.*, 210 (1992) 223.
22. D.N.M. Jones, M. Bycroft, M.J. Lubienski and A.R. Fersht, *FEBS Lett.*, 331 (1993) 165.
23. M.J. Lubienski, M. Bycroft, D.N.M. Jones and A.R. Fersht, *FEBS Lett.*, 332 (1993) 81.
24. M. Overduin, B. Mayer, C.B. Rios, D. Baltimore and D. Cowburn, *Proc. Natl. Acad. Sci. USA*, 89 (1992) 11673.
25. L.G. Riley, F.K. Junius, M.K. Swanton, N.A. Vesper, N.K. Williams, G. F. King and A.S. Weiss, *Eur. J. Biochem.*, 219 (1994) 877.
26. N. Jamin, O. S. Gabrielsen, N. Gilles, P.-N. Lirsac and F. Toma, *Eur. J. Biochem.*, 216 (1993) 147.
27. E.R.P. Zuiderweg and S.W. Fesik, *Biochemistry*, 28 (1989) 2387.
28. N.J. Skelton, M. Akke, J. Kordel, E. Thulin, S. Forsen and W.J. Chazin, *FEBS Lett.*, 303 (1992) 136.
29. J. Anglister, S. Grzesiek, H. Ren, C.B. Klee and A. Bax, *J. Biomol. NMR*, 3 (1993) 121.
30. M. Ikura, L.E. Kay and A. Bax, *Biochemistry*, 29 (1990) 4659.
31. S.M. Roth, D.M. Schneider, L.A. Strobel, M.F. Van Berkum, A.R. Means and A.J. Wand, *Biochemistry*, 31 (1992) 1443.
32. R.A. Venters, T.L. Calderone, L.D. Spicer and C.A. Fierke, *Biochemistry*, 30 (1991) 4491.
33. D.F. Wyss, J.M. Withka, M.H. Knoppers, K.A. Sterne, M.A. Recny and G. Wagner, *Biochemistry*, 32 (1993) 10995.
34. J.C. Madsen, O.W. Sorensen, P. Sorensen and F.M. Poulsen, *J. Biomol. NMR*, 3 (1993) 239.
35. S.W. Fesik, R.T. Gampe, Jr., E.R.P. Zuiderweg, W.E. Kohlbrenner and D. Weigl, *Biochem. Biophys. Res. Commun.*, 159 (1989) 842.
36. G.C. Leo, L.A. Colnago, K.G. Valentine and S.J. Opella, *Biochemistry*, 26 (1987) 854.
37. G.D. Henry and B.D. Sykes, *Biochemistry*, 31 (1992) 5284.
38. F.J.M. van de Ven, J.W.M. van Os, J.M.A. Aelen, S.S. Wymenga, M.L. Remerowski, R.N.H. Konings and C.W. Hilbers, *Biochemistry*, 32 (1993) 8322.
39. R.A. Schiksnis, M.J. Bogusky, P. Tsang and S.J. Opella, *Biochemistry*, 26 (1987) 1373.
40. S. Yajima, Y. Muto, S. Yokoyama, H. Masaki and T. Uozumi, *Biochemistry*, 31 (1992) 5578.
41. P. Neri, R. Meadows, G. Gemmecker, E. Olejniczak, D. Nettesheim, T. Logan, R. Simmer, R. Helfrich, T. Holzman, J. Severin and S. Fesik, *FEBS Lett.*, 294 (1991) 81.

42. K. Wüthrich, C. Spitzfaden, K. Memmert, H. Widmer and G. Wider, *FEBS Lett.*, 285 (1991) 237.
43. R.T. Clubb, V. Thanabal, J. Fejzo, S.B. Ferguson, L. Zydowsky, C.H. Baker, C.T. Walsh and G. Wagner, *Biochemistry*, 32 (1993) 6391.
44. G.M. Smith, L.P. Yu and D.J. Domingues, *Biochemistry*, 26 (1987) 2202.
45. P.R. Gooley, D. Zhao and N.E. MacKenzie, *J. Biomol. NMR*, 1 (1991) 145.
46. B.J. Stockman, M.D. Reily, W.M. Westler, E.L. Ulrich and J.L. Markley, *Biochemistry*, 28 (1989) 230.
47. M.D. Carr, B. Birdsall, T.A. Frenkiel, C.J. Bauer, J. Jimenez-Barbero, V.I. Polshakov, J.E. McCormick, G.C.K. Roberts and J. Feeney, *Biochemistry*, 30 (1991) 6330.
48. B.J. Stockman, N.R. Nirmala, G. Wagner, T.J. Delcamp, M.T. DeYarman and J.H. Freisheim, *Biochemistry*, 31 (1992) 218.
49. R.W. Kriwacki, R.B. Hill, J.M. Flanagan, J.P. Caradonna and J.H. Prestegard, *J. Am. Chem. Soc.*, 115 (1993) 8907.
50. J.W. Peng and G. Wagner, *Biochemistry*, 31 (1992) 8571.
51. D.F. Lowry, R.H. Cool, A.G. Redfield and A. Parmeggiani, *Biochemistry*, 30 (1991) 10872.
52. X. Qian, S.N. Gozani, H. Yoon, C. Jeon, K. Agarwal and M.A. Weiss, *Biochemistry*, 32 (1993) 9944.
53. E. Ab, G.K. Schuurman-Wolters, M.H. Saier, J. Reizer, M. Jacuinod, P. Roepstorff, K. Dijkstra, R.M. Scheek and G.T. Robillard, *Prot. Sci.*, 3 (1994) 282.
54. G.J.A. Kroon, J. Grotzinger, K. Dijkstra, R.M. Scheek and G.T. Robillard, *Prot. Sci.*, 2 (1993) 1331.
55. U. Hommel, T.S. Harvey, P.C. Driscoll and I.D. Campbell, *J. Mol. Biol.*, 227 (1992) 271.
56. T.J. Norwood, D.A. Crawford, M.E. Steventon, P.C. Driscoll and I.D. Campbell, *Biochemistry*, 31 (1992) 628.
57. M.G. Jakoby, K.R. Miller, J.J. Toner, A. Bauman, L. Cheng, E. Li and D.P. Cistola, *Biochemistry*, 32 (1993) 872.
58. T.-M. Chan and J.L. Markley, *J. Am. Chem. Soc.*, 104 (1982) 4010.
59. B.-H. Oh, W.M. Westler, P. Darba and J.L. Markley, *Science (Washington, DC)*, 240 (1988) 908.
60. Y.K. Chae, F. Abildgaard, E.S. Mooberry and J.L. Markley, *Biochemistry*, 33 (1994) 3287.
61. R.D. Guiles, V.J. Basus, I.D. Kuntz and L. Waskell, *Biochemistry*, 31 (1992) 11365.
62. P.R. Gooley, M.S. Caffrey, M.A. Cusanovich and N.E. MacKenzie, *Biochemistry*, 29 (1990) 2278.
63. M. Baron, A.L. Main, P.C. Driscoll, H.J. Mardon, J. Boyd and I.D. Campbell, *Biochemistry*, 31 (1992) 2068.
64. M.K. Rosen, S.W. Michnick, M. Karplus and S.L. Schreiber, *Biochemistry*,

- 30 (1991) 4774.
65. R.X. Xu, D. Nettesheim, E.T. Olejniczak, R. Meadows, G. Gemmecker and S.W. Fesik, *Biopolymers*, 33 (1993) 535.
  66. R.T. Clubb, V. Thanabal, C. Osborne and G. Wagner, *Biochemistry*, 30 (1991) 7718.
  67. B.J. Stockman, A.M. Krezel, J.L. Markley, K.G. Leonhardt and N.A. Straus, *Biochemistry*, 29 (1990) 9600.
  68. M.A. Knauf, F. Lohr, G.P. Curley, P. O'Farrell, S.G. Mayhew, F. Muller and H. Rüterjans, *Eur. J. Biochem.*, 213 (1993) 167.
  69. B.J. Stockman, A. Euvrard, D.A. Kloosterman, T.A. Scahill and R.P. Swenson, *J. Biomol. NMR*, 3 (1993) 133.
  70. T. Mau, J.D. Baleja and G. Wagner, *Protein Sci.*, 1 (1992) 1403.
  71. M. Shirakawa, W.J. Fairbrother, Y. Serikawa, T. Ohkubo, Y. Kyogoku and P.E. Wright, *Biochemistry*, 32 (1993) 2144.
  72. J.G. Omichinski, G.M. Clore, O. Schaad, G. Felsenfeld, C. Trainor, E. Appella, S.J. Stahl and A.M. Gronenborn, *Science (Washington, DC)*, 261 (1993) 438.
  73. P.J.M. Folkers, M. Nilges, R.H.A. Folmer, R.N.H. Konings and C.W. Hilbers, *J. Mol. Biol.*, 236 (1994) 229.
  74. H. Berglund, H. Kovacs, K. Dahlman-Wright, J.A. Gustafsson and T. Hard, *Biochemistry*, 31 (1992) 12001.
  75. W.J. Fairbrother, G.P. Gippert, J. Reizer, M.H. Saier, Jr. and P.E. Wright, *FEBS Lett.*, 296 (1992) 148.
  76. N. Tjandra, V. Simplaceanu, P.F. Cottam and C. Ho, *J. Biomol. NMR*, 2 (1992) 149.
  77. P. Sodano, K.V. Chary, O. Bjornberg, A. Holmgren, B. Kren, J.A. Fuchs and K. Wüthrich, *Eur. J. Biochem.*, 200 (1991) 369.
  78. J.H. Bushweller, A. Holmgren and K. Wüthrich, *Eur. J. Biochem.*, 218 (1993) 327.
  79. T. Zink, A. Ross, D. Ambrosius, R. Rudolph and T.A. Holak, *FEBS Lett.*, 314 (1992) 435.
  80. G.W. Vuister, S.-J. Kim, C. Wu and A. Bax, *Biochemistry*, 33 (1994) 10.
  81. C. Cerf, G. Lippens, S. Muyldermans, A. Segers, V. Ramakrishnan, S.J. Wodak, K. Hallenga and L. Wyns, *Biochemistry*, 32 (1993) 11345.
  82. P.J. Lodi, D.S. Garrett, J. Kuszewski, M.L.-S. Tsang, J.A. Weatherbee, W.J. Leonard, A.M. Gronenborn and G.M. Clore, *Science (Washington, DC)*, 263 (1994) 1762.
  83. D.S. Garrett, P.J. Lodi, Y. Shamoo, K.R. Williams, G.M. Clore and A.M. Gronenborn, *Biochemistry*, 33 (1994) 2852.
  84. M. Wittekind, M. Gorch, M. Friedrichs, G. Dreyfuss and L. Mueller, *Biochemistry*, 31 (1992) 6254.
  85. N.A. van Nuland, A.A. van Dijk, K. Dijkstra, F.H. van Hoesel, R.M. Scheek and G.T. Robillard, *Eur. J. Biochem.*, 203 (1992) 483.

86. R. Boelens, M. Burgering, R.H. Fogh and R. Kaptein, *J. Biomol. NMR*, 4 (1994) 201.
87. S.G. Grdadolnik, M. Eberstadt, G. Gemmecker, H. Kessler, A. Buhr and B. Erni, *Eur. J. Biochem.*, 219 (1994) 945.
88. J.G. Pelton, D.A. Torchia, N.D. Meadow, C.Y. Wong and S. Roseman, *Biochemistry*, 30 (1991) 10043.
89. S. Grzesiek, H. Dobeli, R. Gentz, G. Garotta, A. M. Labhardt and A. Bax, *Biochemistry*, 31 (1992) 8180.
90. B.J. Stockman, T.A. Scahill, M. Roy, E.L. Ulrich, N.A. Strakalaitis, D.P. Brunner, A.W. Yem and M.R. Deibel, Jr., *Biochemistry*, 31 (1992) 5237.
91. D. Marion, P.C. Driscoll, L.E. Kay, P.T. Wingfield, A. Bax, A.M. Gronenborn and G.M. Clore, *Biochemistry*, 28 (1989) 6150.
92. H.R. Mott, P.C. Driscoll, J. Boyd, R.M. Cooke, M.P. Weir and I.D. Campbell, *Biochemistry*, 31 (1992) 7741.
93. C. Redfield, L.J. Smith, J. Boyd, G.M. Lawrence, R.G. Edwards, R.A. Smith and C.M. Dobson, *Biochemistry*, 30 (1991) 11029.
94. R. Powers, D.S. Garrett, C.J. March, E.A. Frieden, A.M. Gronenborn and G.M. Clore, *Biochemistry*, 31 (1992) 4334.
95. G. Lancelot, A. Gervais and J.C. Maurizot, *J. Biomol. Struct. Dyn.*, 9 (1992) 921.
96. P. Leighton and P. Lu, *Biochemistry*, 26 (1987) 7262.
97. H. Matsuo, M. Shirakawa, T. Ohkubo, T. Yamazaki and Y. Kyogoku, *J. Biomol. NMR*, 1 (1991) 191.
98. G.C. Flynn, C.J.M. Beckers, W.A. Baase and F.W. Dahlquist, *Proc. Natl. Acad. Sci. USA*, 90 (1993) 10826.
99. L.P. McIntosh, A.J. Wand, D.F. Lowry, A.G. Redfield and F.W. Dahlquist, *Biochemistry*, 29 (1990) 6341.
100. T. Ohkubo, Y. Taniyama and M. Kikuchi, *J. Biochem. (Tokyo)*, 110 (1991) 1022.
101. C. Freund, A. Ross, A. Plückthun and T.A. Holak, *Biochemistry*, 33 (1994) 3296.
102. G.W. Vuister, R. Boelens, R. Kaptein, R.E. Hurd, B. John and P.C.M. Van Zijl, *J. Am. Chem. Soc.*, 113 (1991) 9688.
103. C. Abeygunawardana, D.J. Weber, D.N. Frick, M.J. Bessman and A.S. Mildvan, *Biochemistry*, 32 (1993) 13071.
104. G.M. Clore, A. Bax, P. Wingfield and A.M. Gronenborn, *FEBS Lett.*, 238 (1988) 17.
105. D.A. Torchia, S.W. Sparks and A. Bax, *Biochemistry*, 28 (1989) 5509.
106. J.F. Wang, A.P. Hinck, S.N. Loh and J.L. Markley, *Biochemistry*, 29 (1990) 102.
107. M. Cox, N. Dekker, R. Boelens, C.P. Verrijzer, P.C. van der Vliet and R. Kaptein, *Biochemistry*, 32 (1993) 6032.
108. E.H. Morita, M. Shirakawa, F. Hayashi, M. Imagawa and Y. Kyogoku,

- FEBS Lett., 321 (1993) 107.
109. A.P. Hinck, W.F. Walkenhorst, W.M. Westler, S. Choe and J.L. Markley, *Protein Eng.*, 6 (1993) 221.
  110. D. Neri, T. Szyperski, G. Otting, H. Senn and K. Wüthrich, *Biochemistry*, 28 (1989) 7510.
  111. T. Szyperski, P. Luginbuhl, G. Otting, P. Guntert and K. Wüthrich, *J. Biomol. NMR*, 3 (1993) 151.
  112. A.R. Peters, N. Dekker, L. van den Berg, R. Boelens, R. Kaptein, A.J. Slotboom and G.H. de Haas, *Biochemistry*, 31 (1992) 10024.
  113. R.J. Gurbriel, C.J. Batie, M. Sivaraja, A.E. True, J.A. Fee, B.M. Hoffman and D.P. Ballou, *Biochemistry*, 28 (1989) 4861.
  114. S. Koyama, H. Yu, D.C. Dalgarno, T.B. Shin, L.D. Zydowsky and S.L. Schreiber, *FEBS Lett.*, 324 (1993) 93.
  115. C.L. Kojiro and J.L. Markley, *FEBS Lett.*, 162 (1983) 52.
  116. W.J. Metzler, K.L. Constantine, M.S. Friedrichs, A.J. Bell, E.G. Ernst, T.B. Lavoie and L. Mueller, *Biochemistry*, 32 (1993) 13818.
  117. S.J. Archer, V.K. Vinson, T.D. Pollard and D.A. Torchia, *Biochemistry*, 32 (1993) 6680.
  118. S. Bagby, T.S. Harvey, L.E. Kay, S.G. Eagle, S. Inouye and M. Ikura, *Biochemistry*, 33 (1994) 2409.
  119. S.L. Campbell-Burk, P.J. Domaille, M.A. Starovasnik, W. Boucher and E.D. Laue, *J. Biomol. NMR*, 2 (1992) 639.
  120. Y. Muto, K. Yamasaki, Y. Ito, S. Yajima, H. Masaki, T. Uozumi, M. Walchli, S. Nishimura, T. Miyazawa and S. Yokoyama, *J. Biomol. NMR*, 3 (1993) 165.
  121. M. Katahira, R.M. Knegt, R. Boelens, D. Eib, J.G. Schilthuis, P.T. van der Saag and R. Kaptein, *Biochemistry*, 31 (1992) 6474.
  122. R. Weisemann, H. Rüterjans and W. Bermel, *J. Biomol. NMR*, 3 (1993) 113.
  123. B.L. Golden, D.W. Hoffman, V. Ramakrishnan and S.W. White, *Biochemistry*, 32 (1993) 12812.
  124. R. Powers, G.M. Clore, A. Bax, D.S. Garrett, S.J. Stahl, P.T. Wingfield and A.M. Gronenborn, *J. Mol. Biol.*, 221 (1991) 1081.
  125. T. Yamazaki, M. Yoshida, S. Kanaya, H. Nakamura and K. Nagayama, *Biochemistry*, 30 (1991) 6036.
  126. R.H. Fogh, D. Schipper, R. Boelens and R. Kaptein, *J. Biomol. NMR*, 4 (1994) 123.
  127. H. Yu, M.K. Rosen and S.L. Schreiber, *FEBS Lett.*, 324 (1993) 87.
  128. H. Torigoe, I. Shimada, M. Waelchli, A. Saito, M. Sato and Y. Arata, *FEBS Lett.*, 269 (1990) 174.
  129. B.A. Lyons, M. Tashiro, L. Cedergren, B. Nilsson and G.T. Montelione, *Biochemistry*, 32 (1993) 7839.
  130. T. Szyperski, S. Scheek, J. Johansson, G. Assman, U. Seedorf and K.

- Wüthrich, FEBS Lett., 335 (1993) 18.
131. G.M. Clore and A.M. Gronenborn, *J. Mol. Biol.*, 223 (1992) 853.
  132. P.R. Gooley, B.A. Johnson, A.I. Marcy, G.C. Cuca, S.P. Salowe, W.K. Hagmann, C.K. Esser and J.P. Springer, *Biochemistry*, 32 (1993) 13098.
  133. S.R. Van Doren, A.V. Kurochkin, Q.-Z. Ye, L.L. Johnson, D.J. Hupe and E.R.P. Zuiderweg, *Biochemistry*, 32 (1993) 13109.
  134. M.L. Remerowski, T. Domke, A. Groenewegen, H.A.M. Pepermans, C.W. Hilbers and F.J.M. van de Ven, *J. Biomol. NMR*, 4 (1994) 257.
  135. F.J. Moy, Y.-C. Li, P. Rauenbuehler, M.E. Winkler, H.A. Scheraga and G.T. Montelione, *Biochemistry*, 32 (1993) 7334.
  136. S.J. Archer, A. Bax, A.B. Roberts, M.B. Sporn, Y. Ogawa, K.A. Piez, J.A. Weatherbee, M.L.-S. Tsang, R. Lucas, B.-L. Zheng, J. Wenker and D.A. Torchia, *Biochemistry*, 32 (1993) 1152.
  137. J.D. Forman-Kay, A.M. Gronenborn, L.E. Kay, P.T. Wingfield and G.M. Clore, *Biochemistry*, 29 (1990) 1566.
  138. D.M. LeMaster and F.M. Richards, *Biochemistry*, 24 (1985) 7263.
  139. K. Chandrasekhar, G. Krause, A. Holmgren and H.J. Dyson, *FEBS Lett.*, 284 (1991) 178.
  140. G.A. Krudy, R.M. Brito, J.A. Putkey and P.R. Rosevear, *Biochemistry*, 31 (1992) 1595.
  141. K.L. Borden, C.J. Bauer, T.A. Frenkiel, P. Beckmann and A.N. Lane, *Eur. J. Biochem.*, 204 (1992) 137.
  142. D.M. Schneider, M.J. Dellwo and A.J. Wand, *Biochemistry*, 31 (1992) 3645.
  143. A.P. Hansen, A.M. Petros, A.P. Mazar, T.M. Pederson, A. Rueter and S.W. Fesik, *Biochemistry*, 31 (1992) 12713.
  144. M.A. Markus, T. Nakayama, P. Matsudaira and G. Wagner, *Prot. Sci.*, 3 (1994) 70.
  145. P. Sorensen and F.M. Poulsen, *J. Biomol. NMR*, 2 (1992) 99.
  146. H.L. Crespi, R.M. Rosenberg and J.J. Katz, *Science (Washington, DC)*, 161 (1968) 795.
  147. J.J. Katz and H.L. Crespi, *Science (Washington, DC)*, 151 (1966) 1187.
  148. J.L. Markley, I. Putter and O. Jardetzky, *Science (Washington, DC)*, 161 (1968) 1249.
  149. L.P. McIntosh and F.W. Dahlquist, *Q. Rev. Bioph.*, 23 (1990) 1.
  150. D.M. LeMaster, *Methods Enzymol.*, 177 (1989) 23.
  151. J. Sambrook, E.F. Fritsch and T. Maniatis, *Molecular Cloning, A Laboratory Manual*, 2nd edn, p. A.3, Cold Spring Harbor Laboratory Press, 1989.
  152. H. Senn, G. Otting and K. Wüthrich, *J. Am. Chem. Soc.*, 109 (1987) 1090.
  153. R.H. Griffey, A.G. Redfield, L.P. McIntosh, T.G. Oas and F.W. Dahlquist, *J. Am. Chem. Soc.*, 108 (1986) 6816.
  154. M. Kainosho and T. Tsuji, *Biochemistry*, 21 (1982) 6273.



155. W.W. Bachovchin and J.D. Roberts, *J. Am. Chem. Soc.*, 100 (1978) 8041.
156. K. Uchida, Y. Miyake and M. Kainosho, *J. Biomol. NMR*, 1 (1991) 49.
157. J.A. Ellman, B.F. Volkman, D. Mendel, P.G. Schultz and D.E. Wemmer, *J. Am. Chem. Soc.*, 114 (1992) 7959.
158. C.J. Noren, S.J. Anthony-Cahill, M.C. Griffith and P.G. Schultz, *Science (Washington, DC)*, 244 (1989) 182.
159. H. Senn, B. Werner, B.A. Messerle, C. Weber, R. Traber and K. Wüthrich, *FEBS Lett.*, 249 (1989) 113.
160. T. Szyperski, D. Neri, B. Leiting, G. Otting and K. Wüthrich, *J. Biomol. NMR*, 2 (1992) 323.
161. J. Rine, *Methods Enzymol.*, 194 (1991) 239.
162. D.T. Moir and L.S. Davidow, *Methods Enzymol.*, 194 (1991) 491.
163. M.S. Neuberger, *Trends Biochem. Sci.*, 10 (1985) 347.
164. M.S. Neuberger, G.T. Williams and R.O. Fox, *Nature (London)*, 314 (1984) 268.
165. L. Riechmann, J. Foote and G. Winter, *J. Mol. Biol.*, 203 (1988) 825.
166. J. Anglister, T. Frey and H.M. McConnell, *Biochemistry*, 23 (1984) 1138.
167. K. Kato, C. Matsunaga, T. Igarashi, H. Kim, A. Odaka, I. Shimada and Y. Arata, *Biochemistry*, 30 (1991) 270.
168. H. Takahashi, A. Odaka, S. Kawaminami, C. Matsunaga, K. Kato, I. Shimada and Y. Arata, *Biochemistry*, 30 (1991) 6611.
169. H. Takahashi, E. Suzuki, I. Shimada and Y. Arata, *Biochemistry*, 31 (1992) 2464.
170. K.-M. Lo and S.D. Gillies, *Bioch. Bioph. Acta*, 1088 (1991) 217.
171. L.E. Gentry, N.R. Webb, G.J. Lim, A.M. Brunner, J.E. Ranchalis, D.R. Twardzik, M.N. Lioubin, H. Marquardt and A.F. Purchio, *Mol. Cell. Biol.*, 7 (1987) 3418.
172. K. Shon, P. Schrader, S. Opella, J. Richards and J. Tomich, *Frontiers of NMR in Molecular Biology*, pp. 109–118, Liss, Inc., New York, 1989.
173. J.L. Markley and I. Kato, *Biochemistry*, 14 (1975) 3234.
174. M.W. Baillargeon, M. Laskowski, Jr., D.E. Neves, M.A. Porubcan, R.E. Santini and J.L. Markley, *Biochemistry*, 19 (1980) 5703.
175. R.H. Griffey and A.G. Redfield, *Q. Rev. Biophys.*, 19 (1987) 51.
176. S.W. Fesik, H.L. Eaton, E.T. Olejniczak, E.R.P. Zuiderweg, L.P. McIntosh and F.W. Dahlquist, *J. Am. Chem. Soc.*, 112 (1990) 886.
177. G.M. Clore, A. Bax, P.C. Driscoll, P.T. Wingfield and A.M. Gronenborn, *Biochemistry*, 29 (1990) 8172.
178. G.M. Clore, L.E. Kay, A. Bax and A.M. Gronenborn, *Biochemistry*, 30 (1991) 12.
179. G. Wagner, *Prog. NMR Spectr.*, 22 (1990) 101.
180. A.C. de Dios, J.G. Pearson and E. Oldfield, *Science (Washington, DC)*, 260 (1993) 1491.

181. S. Grzesiek, J. Anglister, H. Ren and A. Bax, *J. Am. Chem. Soc.*, 115 (1993) 4369.
182. G. Wagner, *J. Biomol. NMR*, 3 (1993) 375.
183. S.J. Matthews and R.J. Leatherbarrow, *J. Biomol. NMR*, 3 (1993) 597.

## Chapter 4

# Protein Structure Determination from NMR Data

PAUL L. WEBER

### 1. INTRODUCTION

Significant progress has been made in the past few years for determining protein structures from nD NMR data. Compared with the initial NMR structures of ten years ago, the structures being determined today are of much larger molecules, are determined to higher precision and accuracy, and are being determined in much less time. NMR is seen now as being quite complementary to X-ray diffraction methods, and NMR structures are also being accepted by the general scientific community as being accurate structures, accounting for about 20% of the new structures determined during the past few years (the majority still being determined by X-ray crystallographic methods).

Many developments have been required for these advances since the earliest studies in the 1970s. One of the most important is certainly the availability of high field spectrometers (500–750 MHz) and multi-dimensional (nD) techniques. Of obvious importance is the development of multichannel spectrometer hardware and indirect detection techniques. Other advances are due to continuing improvements in computer hardware and software, plus advances in techniques for calculating protein structures using NMR data. Besides the technology advances, there has been considerable growth in the number of people using NMR techniques to determine protein structure, as has been seen by the greatly increased attendance and number of posters and talks describing structure determinations at recent biological NMR conferences.

Despite this growth, the field is still in its relative infancy. This chapter has the dual purpose of describing current problems as well as future potentials. Some of the common methods used will be discussed, and the relative strengths and weaknesses of each method will be presented.

This chapter is not, by any means, meant to be a review of the many structures that have been calculated by NMR. It is intended to be a survey of the approaches used to calculate structures from NMR data, often examining details of the process that some might normally take for granted. However, the nature of NMR-based structure determination is such that nothing can be taken for granted, lest assumptions made come back to haunt the user in the end. And they have.

One goal of this chapter is to highlight how the different steps for calculating structures can best be used to enhance the success of subsequent steps. While the steps are presented as a simple linear list of tasks, structure calculations are rarely performed this way. Usually there is considerable iteration between different steps of the process. The strategies one uses to iterate between the different stages can strongly influence the success of the structure determination. A second goal is to examine the many different options that are available. Those routines that have clearly gone out of favor will be mentioned, so that such methods may be approached with due caution. The final goal of this chapter is to bring attention to areas of either exceptional promise or importance.

## ***1.1. An Overview of NMR-based protein structure determination***

### ***1.1.1. Determine resonance assignments***

It is possible to obtain limited qualitative information about a protein structure without any resonance assignments. One can often tell from proton chemical shift dispersion and the relative ratios of different NOE types something about the relative proportions of secondary structure elements in a protein. For example, the presence of many strong NOEs between alpha protons is indicative of antiparallel beta sheet structure. Nonetheless, complete or nearly complete resonance assignments are required for any specific structural information to be derived. Therefore the first step in any structure calculation is to obtain as complete and as accurate a list of resonance assignments as possible. Typically the *proton* assignments are the most necessary (for obtaining proton-proton

distances from NOE spectra). However, the increasing use of heteronuclear J-couplings, as well as the rapidly growing use of heteronuclear experiments as a means of obtaining resonance assignments in larger proteins requires assignments to be obtained for proton, carbon and nitrogen nuclei. The addition of nitrogen and carbon assignments is also key to determining more structural constraints as the unambiguous identification of NOEs becomes straightforward, even in molecules approaching the theoretical limit of NMR-based structure determination in solution, namely ca. 30,000 MW.

These topics are covered in more detail in Section 2.

#### *1.1.2. Generate a list of experimentally-derived geometric constraints*

The most common constraints obtained are distances derived from the relationship of NOE intensity to internuclear separation, and torsion angles derived from couplings using the well-known Karplus relationships. It is also possible to derive hydrogen bond constraints from NMR data, and these are often included as constraints when calculating structures. Constraints should be as precise and accurate as reasonable, given the accuracy of the data and the motions present in the molecule; constraints also are usually generated based on unambiguous atom assignments for all resonances involved, although recent developments in the use of ambiguous constraints seem promising.

Constraint generation is covered in Section 2.

#### *1.1.3. Generate a starting conformation*

The choice of a starting conformation is critical for three reasons:

1. "Operator bias" must be avoided, lest it influence the final structures;
2. Sampling of accessible conformation space is desired in order to determine the full range of conformers consistent with the data;
3. Depending on how structure optimization is performed, significant barriers to folding may be present, reducing computational throughput or speed.

Starting structures are usually randomly-chosen structures, or model structures from X-ray crystallography, model building, or distance geometry. These are covered more in Section 4.1.

#### *1.1.4. Optimize the conformation to best fit the experimental constraints*

Next the conformation of the starting structure is modified until the constraint errors are minimized. The optimization routines usually

employed are various minimizers, molecular dynamics and dynamical simulated annealing. Minimization may be performed in Cartesian coordinate space (in three or more dimensions), distance space, or torsion space. These options also may use different optimization protocols. While it is not possible to exhaustively review all of these options, those that appear often in the literature, or those with much promise will be described in this chapter. These options are described in Sections 4.2–4.4.

#### 1.1.5. *Analyze the resulting structure for agreement with the data*

The last step is to compare calculated structures to the applied constraints to determine the success of the optimization stage. Also, structures from repeated calculations are compared to judge the level of structural convergence to determine if adequately determined structures have been obtained. This last stage should also include a comparison of the experimental spectra with theoretical spectra calculated from the derived structures, but this is not always performed. Initial structures are also used to further identify unambiguous constraints from the experimental data. Finally, all of these steps are repeated until satisfactory results are obtained.

Analysis options are described in Section 5.

## 2. DETERMINING RESONANCE ASSIGNMENTS

Structural information is obtained from features present in nD NMR cross peaks, such as peak intensity, J-couplings between individual peaks in a multiplet, or at the simplest level, from the peak chemical shifts. In all cases, the information present in a peak is not generally useful until the identity of the atoms that give rise to this peak are known. The correlation of atom identities to resonance positions (i.e., chemical shift) is known as *resonance assignments*.

Several strategies for obtaining assignments have been published, but all strategies are based upon comparing observed patterns of chemical shifts and couplings (either J-couplings or NOEs) with the expected patterns for the protein, given the two-dimensional atomic structure. For example, DQF-COSY data of a protein might reveal a spin system composed of an amide proton, an alpha proton, and a methyl proton; this would be *identified* as an Alanine residue. If NOESY data showed this Alanine's amide proton to be close to both the amide, alpha and beta

proton of a Valine residue, and if the only occurrence of a Val–Ala dipeptide occurred at positions 43–44 of the protein, then one could *assign* the resonances in these Valine and Alanine spin systems to V43 and A44. This method is known as the *sequential assignment method*. Instead of relying on NOEs to link the Valine and Alanine, one might have acquired experiments that illustrated heteronuclear J-couplings through the peptide bonds. Such methods will be described later in this section.

### ***2.1. Sequence-specific proton resonance assignments in proteins***

The sequential assignment methodology was the first method that allowed complete proton resonance assignments to be obtained for protein NMR spectra. The method is still very much in use, and a brief summary of it follows. Those wishing a fuller explanation of sequential assignments, as well as an excellent (albeit somewhat outdated) description of protein NMR methods are referred to the book *NMR of Proteins and Nucleic Acids* by Kurt Wüthrich, in whose lab these methods were developed.

Prior to performing the sequential assignments, one must first identify most or all of the coupled proton spin systems of the individual amino acids. This is done through the use of CMQF-COSY, MQ-COSY, TOCSY and/or RELAYED-COSY types of experiments. Identification of the spin systems involves matching patterns of coupled cross peaks with expected patterns. Usually the (proton) chemical shifts of the resonances are used to further discriminate between different spin systems, but complete reliance on chemical shift information may be dangerous.

The principle of the sequential assignment method relies on the observation that the probability is very high that the amide proton of residue  $i$  is close to the amide, alpha and/or beta protons in the preceding residue ( $i-1$ ), and that the probability is very low that the amide of residue  $i$  is close to the amide, alpha and beta protons of any other residue than the preceding one (see Fig. 1). Thus, NOESY spectra are examined to determine all such nearest-neighbor NOEs.

Following identification of spin systems and the sequential NOEs to link them, one compares the derived peptide sequence information to the published or otherwise known protein sequence, as mentioned in the example on the previous page. Unique peptide sequence matches will yield sequence-specific resonance assignments.

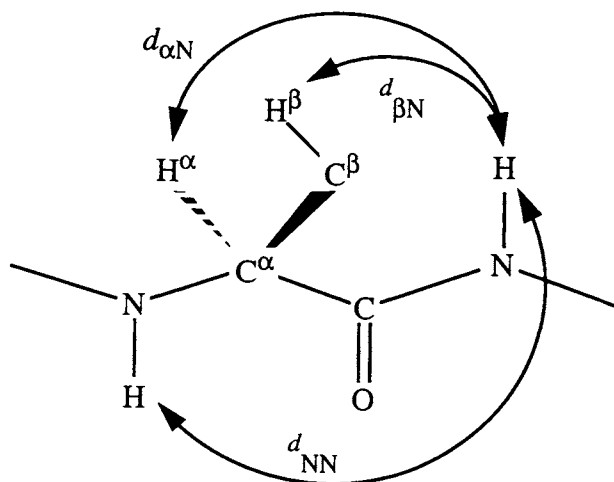


Fig. 1. The three major sequential backbone NOE connectivities.

In practice one often cannot identify complete amino acid spin systems prior to making all of the sequential connectivities. Sequential assignments are often obtained in an iterative fashion, where knowledge of some unambiguous sequential assignments is used to eliminate ambiguous interpretations of unassigned resonances. Also, in proteins with considerable secondary structure, one may make good use of additional backbone NOEs that help define helical and sheet structures. For example, one may be able to use NOEs that occur across an antiparallel beta sheet to make or confirm sequence-specific assignments, when no sequential NOEs are otherwise observed (perhaps due to overlap of alpha proton and solvent resonances).

Such a strategy may even be taken to the extreme, where the patterns of NOEs observed in secondary structure may be used as the primary discriminating means to obtain resonance assignments. This strategy has been formalized and named the "Main Chain Directed" (MCD) approach [2], following the emphasis on use of the backbone NOEs, and de-emphasis on making complete amino acid spin system identifications prior to sequence-specific assignments. This approach is also suggested to be more amenable to automation of assignments, but it has a weakness in that it is only useful for structures with regular secondary structure. Fortunately, it is usually the case that regular secondary structure is present.



Sequential assignments of proteins is limited in practice to molecules with less than 100–120 residues, due to spectral overlap. The use of only proton spectra to determine assignments is becoming less common, but is often the only method available when the samples of interest cannot be isotopically enriched (see below).

## ***2.2. Heteronuclear resonance assignments***

An area of rapidly-expanding applications is in heteronuclear NMR of proteins, using samples enriched in  $^{13}\text{C}$  and  $^{15}\text{N}$  isotopes. Heteronuclear spectra are used for several purposes:

1. obtaining resonance assignments without reliance upon NOE connectivities;
2. obtaining information on dynamic processes;
3. obtaining additional geometric constraints; and
4. as a general means for filtering information present in crowded protein spectra.

Some of these techniques have been discussed elsewhere in this volume, and no further introduction to these techniques will be presented here.

Of notable importance to the scope of this chapter is the growing suite of 3D  $^1\text{H}$ – $^{13}\text{C}$ – $^{15}\text{N}$  experiments developed in several laboratories [3,4]. These experiments are designed to filter out all but selected couplings between backbone atoms, giving rise to uniquely identifiable backbone fragments. Overlapping coupling patterns are used to generate overlapping fragments, from which the entire protein backbone may be followed in a virtually unambiguous manner.

The experiments performed to obtain proton, nitrogen and carbon assignments are still in a state of rapid development at the time of writing this chapter. The strategy for obtaining assignments, however, seems very common and is not likely to change. That is, the experiments performed are designed to observe couplings between very specific sets of backbone atoms, and that comparison of two or more such data sets will allow the user to determine the resonance assignments.

Extensive data to complete side chain resonance assignments may next be obtained using additional experiments that utilize J-couplings between carbon and proton nuclei. For example, HCCH experiments have been very successful in spreading the chemical shift information into three or four dimensions, making unambiguous assignments and NOE identifications possible even in the crowded aliphatic region of the proton spectrum [5–7].

The advantages of heteronuclear correlation experiments to obtain resonance assignments are numerous. First, spreading the information into three-dimensions is clearly advantageous versus 2D approaches. Second, the experiments used are highly selective for specific three-atom correlations, and are often for three different nuclei ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ). This greatly simplifies the task of identifying the atoms involved in the correlations. Third, it places much less reliance on NOE connectivities, which can easily become ambiguous in larger proteins. Last and perhaps most important, this approach can be automated readily. But an obvious and serious disadvantage of these methods is that they require samples enriched in  $^{15}\text{N}$  and  $^{13}\text{C}$ , which will not be available for proteins that are either not cloned or expressed in systems that allow isotope enrichment. In addition, the heteronuclear couplings become difficult to observe as the protein size increases above 30,000 MW.

### ***2.3. Stereospecific proton resonance assignments***

It is usually the case that two methylene protons on a carbon adjacent to a chiral center will have different chemical shifts. An analogous case is found for the methyl groups of Valine and Leucine residues. Although they are otherwise chemically equivalent, the different chemical shifts means that these two protons can be distinguished from each other in the proton NMR spectrum. The problem comes in determining which signal corresponds to the pro-S proton, and which is the pro-R proton. This is known as the problem of stereospecific resonance assignments.

#### ***2.3.1. Pseudoatom corrections***

Pseudoatoms may be used when stereospecific assignments are not available [1]. In this approach, a pseudoatom is defined to lie midway between the stereo-related protons (or methyls). NOEs to either stereo-related resonance are used to define constraints to the pseudoatom, where the upper bound is increased by at least the distance between the pseudoatom and the actual atom involved. Use of pseudoatoms is a compromise versus not using the NOE information at all, as they will reduce the precision of calculated structures due to the weakening (widening) of the constraint information.

#### ***2.3.2. Direct determination of stereospecific assignments***

The preferred solution to this problem is to use the available constraints and geometric information to see if they unambiguously define

the pro-chirality. This can be done using intra-residue and sequential NOEs [8], homonuclear coupling constants [9], and/or heteronuclear coupling constants [10]. These approaches are often automated using grid search routines [9,11] to ensure an unbiased, systematic approach to obtaining the assignments prior to actual structure calculations.

Another method of obtaining stereo-assignments involves incorporation of selectively deuterated or fractional  $^{13}\text{C}$  labelling [12,13]. In contrast to the previous methods, these are laborious, costly, and limited in applicability, but have the advantage of yielding perfect confidence in the stereo-assignments.

### *2.3.3. Indirect determination of stereospecific assignments*

A modest drawback of the “direct” methods is that they do not take advantage of the effect of long-range distance constraints (i.e., from NOEs between amino acids well-separated in the protein sequence) or the additional topological limitations placed by the global fold of the molecule. This information is available for making stereo-assignments after initial structures have been calculated, where the methylene protons (or methyls) are analyzed to compare the calculated position (pro-R or pro-S) with an initial “guess” stereo-assignment, and a reversed “guess” stereo-assignment. Repeating this for several structures gives one a measurable statistical confidence in calculated stereo-assignments. This approach has been reported for the case where prochiral groups were allowed to “float”, or freely invert between pro-R and pro-S forms [14], and for the case where the upper bounds violations for the input “guess” stereo-assignments were compared to reversed guess assignments [15]. This latter case was also automated using a computer program (HABAS, Ref. [11]) to determine stereo-assignments prior to structure calculations, as well as using a second program (GLOMSA) to determine assignments after structure calculations were done.

## ***2.4. Identification of secondary structure from sequential assignments***

The patterns of NOEs used in making assignments (assuming that homonuclear methods are being used) will indicate the secondary structure elements present in the molecule, especially when used in conjunction with  $J_{\text{N}\alpha}$  couplings and amide exchange data [1]. This information can be quite useful for the scope of some applications, and is often much more interesting than the rather dry results of resonance assignments.

It is included in this section on assignments because the identification of regular secondary structure will both increase the confidence in the assignments, as well as provide means for extending assignments using non-sequential backbone NOEs.

To determine secondary structure with a reasonable level of confidence in the results, one must use more than the  $d_{\alpha N}$ ,  $d_{\beta N}$  and  $d_{NN}$  sequential assignments. For example,  $d_{NN}$  connectivities alone do not distinguish an alpha helix from a  $3_{10}$  helix. To distinguish these two classes of substructure, one must rely on the differences of NOE patterns given the different repeat length. (The length of the helix might also be used to distinguish the two, as long runs of  $3_{10}$  structure are unknown).

To continue this example, additional support of a helical structure would be the presence of slowly-exchanging amide protons (except for the first four residues). However, slowly exchanging amides are not always observable, due to the possibility of protein instability at the pH levels that are best for observing amide signals.

To date, no information is available to indicate secondary structure when assignments are obtained using heteronuclear correlation methods [4]. This is not really a shortcoming of such methods, as most protein NMR studies that include obtaining complete resonance assignments continue on to the step of generating three-dimensional structures using constraint information.

### 3. GENERATING EXPERIMENTAL CONSTRAINTS

Nuclear magnetic resonance spectra may contain a lot of information about a molecule's structure, but rarely can that information be directly translated into an *exact* geometric constraint. This is due mostly to using inaccurate or incomplete models for calibrating the geometric information to the experimental values. The best known example of this is the problem of calibrating distance to NOE intensity. Most often one attempts to calibrate a distance based on an isolated two-spin model, even though this is clearly not an accurate representation of a protein. The errors encountered when using this approach can be quite substantial (as much as 30%).

Fortunately, proteins are one of the few cases where nature provides enough information to overcome her otherwise mischievous behavior. It has been shown that quite reasonable structures can be calculated

when many relatively inaccurate constraints are used [16]. Furthermore, as protein size increases, the redundancy in the constraints increases, hence structures of larger proteins (100 to 200 residues) may be determined to higher overall accuracy than smaller proteins (less than 60 residues). For a good quality structure, about 10–15 NOE constraints per residue will be required, although low resolution structures using as few as 2–5 NOEs per residue are possible. Such low resolution structures may be invaluable for making additional unambiguous NOE identifications. For example, if chemical shift overlap resulted in a given NOE having two possible identities (“A” to “B” or “A” to “C”), examination of the two corresponding distances in the initial low resolution trial structures often will show that one distance is always less than 5–7 Å, whereas the other distance is always greater than 8–10 Å.

Five types of constraints will be covered in this section: covalent constraints, non-bonded distance constraints, torsion angle constraints, hydrogen bond constraints (which may include both a distance and angle term), and chemical shift constraints.

### **3.1. Covalent constraints**

These are constraints that one often has prior to even starting the process of assigning the protein. Covalent constraints essentially are the bond lengths and bond angles in the molecule. However, this assumes that the protein sequence is known *correctly*, which is not always the case for proteins obtained from natural sources. Many examples are known where published peptide sequences did not match the sequences deduced from NMR analysis [1, and references therein].

In addition to confirming or correcting the amino acid sequence, NMR constraints may also be obtained that indicate disulfide bonding patterns in the protein [16]. In most cases disulfide bonds are inferred from analysis of NOEs between the beta protons of the cysteine residues. For metal-bonding proteins, metal coordination constraints may be determined experimentally, for example from  $^1\text{H}$ – $^{113}\text{Cd}$  couplings [18–20]. These constraints may have a profound effect upon the folding of the molecule, or in limiting the conformational flexibility of the molecule. Such information may not always be available from non-NMR sources, and may provide additional important information about the structure and function of the molecule.

### 3.2. Distance constraints

NOE-derived distance constraints are by far the most common constraint used in NMR-based structure calculations. NOEs contain valuable information on both the structure and dynamics of the molecule, and there are several means currently in use to extract quantitative distance information from NOE measurements. All of these methods are dependent upon the ability to accurately measure the NOE volumes (which in practice is not trivial) and to take into account the overall tumbling of the molecule as well as its internal motions (which is neither straightforward nor trivial).

#### 3.2.1. Pitfalls in measuring NOEs

Before describing how one measures NOEs and obtains distance constraints, I will first discuss why it really cannot be done as well as we would like, hope, or think.

First is the problem with measuring volume integrals, especially in noisy, under-resolved and crowded spectra. The problems associated with low S/N and overlapped peaks are fairly obvious, but little consideration is ever given (or at least mentioned in the literature) as to how the actual integration is performed. Since the data are digitized, the software performing the integration needs to determine which points will be used in performing the integration. This decision is two-headed: (1) is a rectangular area to be integrated, or an ellipsoidal area; and (2) will the data be only integrated for discrete points, or will an interpolation be performed to estimate the values that fall between the digitized data point and the peak edge?

The next major pitfall is motion. The cross relaxation rate of the system depends upon intermolecular distances, the overall tumbling rate(s), and the effects of internal, or localized motions. Most proteins tend towards isotropic tumbling, but the effects of anisotropic tumbling can be very severe [21]. The effects of local motion can also affect the calculated cross-relaxation rate between 5 and 50%, depending on the rate of local motion compared with the rate of overall molecular tumbling. Since many local motions, such as methyl rotations, are not very well understood, any comments about the accuracy of one's structures should be made with the knowledge that we really do not know what the accuracy of any cross relaxation measurement really is.

The last pitfall is in what you do not know, or take for granted. If your cross relaxation model does not take into account the effects of, say,

paramagnetic impurities in your sample, then the calculated cross relaxation rates can become quite meaningless.

With these caveats stated, I will now present how we proceed and either include them in structure refinements, or ignore them.

### 3.2.2. NOE volume/intensity measurements

To obtain NOE distance constraints, one could reference the NOE intensity of a peak to the intensity of an NOE between two protons whose interproton distance is known, using the familiar  $r^6$  dependence of the NOE intensity to the interproton separation:

$$\frac{\text{NOE}_{\text{unknown}}}{\text{NOE}_{\text{reference}}} = \frac{r_{\text{reference}}^6}{r_{\text{unknown}}^6}$$

This method is subject to error due to variations in the effective correlation times between different spin pairs, time-averaging of the distances, as well as errors due to the deviation of the NOE intensity from a linear build-up rate. Nonetheless, the  $r_{ij}^{-6}$  dependence readily allows *approximate* distance constraints to be obtained.

When calibrating distances in this manner, the choice of reference distance is important. This distance should be between two protons that are (1) obviously separated by a known distance; (2) not subject to greater motional averaging than most interproton pairs being calibrated; (3) not subject to significant strong coupling effects. In addition, it should be obvious that NOEs to methyl protons will also need to be scaled by one-third due to the three protons involved [22]. It is also known that errors in NOE estimation will lead to calibrated distances that are skewed towards the reference distance [23]. Personally, when using this approach I favor the use of a strong  $d_{\alpha N}$  NOE in a beta-sheet region for a reference distance. While it is not geometrically rigid, it varies only between 2.18 and 2.24 Å when the  $\phi$  torsion varies  $\pm 20^\circ$  about the *trans* configuration, and it is not subject to the problems of motion or strong coupling.

The strong distance dependence helps to buffer against errors in NOE measurements. This is most welcome, as measuring an accurate and precise peak volume from a NOESY spectrum can be very tricky. One common problem is that most apodization functions used during data processing to improve resolution can alter the integral values noticeably. On the other hand, poor resolution can lead to problems in integral measurement, especially in crowded spectra or in the presence of strong artifacts like baseline tilts,  $t_1$  streaks, etc. Some means might

be used to derive better integrals include lineshape fitting routines [24] and “pie slice” reconstructions [25]. These are described below.

Comparison of symmetry-related cross peaks may be helpful (especially in spectra with different  $t_1$  and  $t_2$  resolution), although one is presented with the problem of how to treat symmetry-related NOEs that have different volume integrals: does one take the average of the two peaks, just use the strongest value, or just use the weakest value? One might even set a lower distance bound using the stronger NOE, and the upper distance bound using the weaker NOE. In some cases, if no symmetry-related NOE is observed for a particular peak, the peak might be considered to be an artifact.

Lineshape fitting routines involve attempting to fit the experimental data with a model lineshape in each dimension. This method allows for each peak to be fitted separately, and a different lineshape model used for each resonance in each dimension. Integration of the peak volume can also correct for baseline contributions, since perfectly flat baselines are difficult to achieve in practice using conventional FT routines. An added benefit of lineshape fitting procedures is that a goodness-of-fit parameter can be obtained, allowing one to assess the confidence in each NOE volume.

A limitation of lineshape analyses that has been reported is usually that the lineshapes are too simplified, namely a single Gaussian is sometimes used when typically cross peaks can have significant fine structure.

The “pie-slice” reconstructions are the most simple, where a segment of a peak that is not overlapped (perhaps a half, or quarter) is specified, integrated and then multiplied by a scaling factor to estimate the entire peak volume (e.g., 2 if a half-peak is used). This approach has the benefit of being the easiest to implement, and can be used to overcome a variety of artifacts. It can, however, only be used with any accuracy when the peak overlap is slight.

### *3.2.3. Relaxation matrix approaches to determining distance constraints*

A major problem with the “two spin” approximation for scaling NOE intensities or volumes to distances is that the effects of neighboring spins and their relaxation are not considered. It is well known that this results in inaccurate distance calibration, and can be quite considerable if the distances are obtained from NOESY experiments acquired with long mixing times. An alternative method is to solve the Bloch equations for the systems relaxation, and thus obtain distances with better accuracy than with the two-spin approach.



The use of relaxation matrix methods is now widespread, with the programs MARDIGRAS, IRMA and MORASS all readily available. It is outside the scope of this chapter to describe these methods in detail, but how their use affects structure quality is offered. First, all of these methods require the accurate measurement of NOE volumes. This is still problematic for two reasons: (a) to obtain theoretically accurate NOE measurements, one would need to record their NOESY mixing times where sufficient relaxation delays are included, and this is almost never done; (b) the NOE cross peak volumes must be accurately measured. Simple integration routines are susceptible to many errors, such as not interpolating data intensities at the peak edges, or not correcting the spectrum's baseline prior to integration.

The program MARDIGRAS used in my group has one feature worth noting, namely that one can obtain a reasonable estimate of the inaccuracy present in the data by comparing the intensities of symmetry related peaks. The program will return distance constraints where the lower and upper bounds are adjusted to take into account the known inaccuracies. Greater inaccuracy will lead to wider bounds constraints, as expected.

#### *3.2.4. NOE-derived constraint precision*

How distance constraints are used varies greatly between users, although four general classes of distance constraints may be identified:

1. An equilibrium distance for use in a pseudoenergy potential function [26];
2. Lower and upper bounds, determined from an exact distance  $\pm$  some error [14];
3. Lower and upper bounds, using "strong-medium-weak" ranges [1];
4. Set upper bound only, e.g., to 5 Å [27].

Each method has its own advantages and disadvantages, as well as its own followers and detractors. The use of relatively wide bounds is perhaps the most common in proteins, and especially in larger proteins where sufficient constraints may be identified and where a well-folded structure allows the use of less precise, but multiply overlapping and topologically restrictive constraints [16]. Such loose constraints also reduce the susceptibility to errors due to the effects of spin diffusion, which is a major relaxation phenomenon in proteins. Relatively imprecise distance constraints also reduce the need to obtain precise NOE peaks volumes, which can be difficult (as described in the previous

section). On the other hand, narrower constraints can lead to better converged structures, but this will be meaningful only if the distances are accurate.

Distance constraints may be refined through the use of relaxation matrix methods that account for spin diffusion in the protein, and in some cases for the differences in relaxation for methyl groups [28]. Such a procedure reduces the need to obtain precise distance constraints initially; instead, relatively imprecise constraints are used as a starting point, and the distance constraints are refined as the structure is refined. It is likely that iterative methods such as this will become more standard for both refining distance constraint estimates and structures.

### 3.2.5. Use of non-NOE information

One has the choice to use “non-NOE” information, where the lack of an NOE is taken to indicate that two protons are *further apart* than a given threshold distance. Given the potential for NOEs to be not observed in spectra due to experimental artifacts, spin diffusion, or sample problems, non-NOE information should be used only with great caution. Non-NOE information, however, can be quite valuable if it can be shown to be reliable.

What might be considered reliable non-NOE information? To answer this, first examine some of the reasons that a short interproton distance might not be observable in a NOESY spectrum. First, there are spectral artifacts due to the experimental setup or due to improper processing. Examples include “bleaching” due to presaturation, or  $\omega_1$  streaks due to incorrect FT treatment of the first data point [29]. Therefore one would *not* consider non-NOE information from a part of the spectrum where bleaching or streaking is strong. Another reason for loss of NOE intensity could be due to strong localized relaxation, such as could be present from bound paramagnetic ions, or in very flexible regions, or in regions where spin diffusion could weaken NOEs. In these cases one would not consider non-NOE information for those atoms that *never* appear in NOEs, or are consistently weak. But if, for example, an NOE is observed between amide proton “A” and the beta proton “B1” of another residue, but not the geminal-related “B2” proton, then this would be compelling evidence that A and B2 are at least further apart than A and B1, providing that B2 is in a region of the spectrum devoid of artifacts.

Non-NOE information is also important if one is refining structures with a force field that includes Lennard–Jones attractive terms, and/or electrostatic terms. Such forces could lead to artificial contacts forming

between atoms, where experimental evidence “clearly” shows no such contacts exist. Non-NOE information is used in these cases to prevent such contacts from forming.

### 3.3. *Torsion angle constraints*

For high quality structures, it is important to obtain as many of the  $\phi$  and  $\chi_1$  torsion constraints as possible prior to starting the structure calculations.

The  $\phi$  and  $\chi_1$  torsion angles may be calculated from J-couplings using the familiar Karplus equations. These may be used directly in programs that have penalty functions based on the torsion values, or may be converted to distance constraints for those programs that do not. Torsion constraints are much more valuable than the converted distance constraints, as they are able to, in many cases, better restrict the conformational space of the system. For example, a distance constraint between an alpha and beta proton set to  $2.5 \text{ \AA} \pm 10\%$  is consistent with torsion angles from  $-100$  to  $+100^\circ$ . While this is more than half of the accessible conformation space, the distance constraint used might even be considered overly precise by some.

Accurate J-couplings usually cannot be extracted easily and directly from protein DQF-COSY spectra, due to the linewidths generally found in proteins. These will often cause the apparent J-coupling to be larger than the actual coupling [30], especially when the actual J-coupling is weak ( $<6 \text{ Hz}$ ). To overcome this problem, either spectral simulations [9] or line fitting [31,32] of peaks can be used to obtain coupling constants with relatively good precision.

In addition to using DQF-COSY data to measure J-couplings, E-COSY type experiments [33,34] may be used to overcome the problems of broad lines. These experiments are useful for identifying both active and passive couplings. Another method that is useful for proteins that have been isotopically enriched is to measure sequential  $^{15}\text{N}^1\text{H}^\alpha$  (for  $\phi$  torsions) and  $^{15}\text{N}^1\text{H}^\beta$  couplings (for determining both  $\chi_1$  torsions and stereo-assignments) [35]. Heteronuclear correlation experiments using  $^{13}\text{C}$  are also useful for measuring  $\text{C}^\alpha\text{C}^\beta$  torsion angles [36].

The  $J_{\text{H}\alpha\text{H}\beta}$  values for serine and threonine also need to be corrected for the high electronegativity of the oxygen [37]; failure to do so will result in J values that are about 10% too high. Once accurate J-couplings have been measured, torsion angles may be determined using the Karplus relationship [38]:

$$^3J(\theta) = A \cos^2\theta + B \cos \theta + C$$

The values of the three constants have been estimated for both  $J_{\text{NH}\alpha}$  and  $J_{\text{H}\alpha\text{H}\beta}$  torsions [39–41] among others.

Torsion constraints, like distance constraints, are usually expressed as a range of acceptable torsions. The width of these ranges may be derived from the limit in precision for measuring J-couplings (often  $\pm 1$  Hz), or the width may be something relatively arbitrary, such as a  $\pm 30^\circ$  variance. Setting torsion constraints to narrow ranges will likely limit the conformations that can be calculated, but care must be taken to show that this apparent convergence is merited from the data.

A problem with torsion constraints is that there is not a unique match between a coupling constant  $^3J$  and the dihedral angle  $\theta$ , as there are two maxima and minima for the Karplus relationship. Nonetheless, when the coupling information is combined with distance constraints and/or additional coupling information, one can narrow the possible torsion angles to a single value or limited range [9,10]. One might even choose not to calculate a torsion constraint at all, but rather to minimize based on the calculated J-couplings:

$$V_J = \sum_{\text{torsion}} K_J (^3J(\theta) - ^3J(\theta_0))^2$$

where J is evaluated for the current values of the constrained torsions, and compared to the constrained value. This way the different torsion ranges that satisfy the experimental J values will automatically be allowed. If multiple coupling and/or distance constraints further restrict the allowed torsion range, then they will also fully affect the structure calculation, without any need for *a priori* determination of the allowed torsion range for any given coupling value.

Another problem in obtaining good torsion or J restraints is that molecular motions will result in averaging of the torsion values. The approach sometimes used for time-averaged *distance* constraints (see Section 4.3.4) can be extended to the treatment of torsion constraints, modifying the previous equation to:

$$V_J = \sum_{\text{torsion}} K_J (\langle ^3J(\theta) \rangle - ^3J(\theta_0))^2$$

### **3.4. Hydrogen bond constraints**

Hydrogen bonds can often be inferred from proton spectra if several indicators are present. First is the presence of amide protons protected from exchange, although this alone is not unambiguous evidence (amide exchange can be restricted due to steric inaccessibility). Nonetheless, the presence of slow amide exchange along with regular patterns of secondary structure deduced from sequential NOEs is a very good indicator of hydrogen bonding patterns. More evidence for hydrogen bonds can be gained from the temperature coefficients of the amide chemical shifts [42].

Another indication of hydrogen bonding is the chemical shifts of amide protons [43], or the carbonyl carbons [44], although again this is not sufficient evidence by itself. In ubiquitin, a hydrogen bond between a tyrosine hydroxyl group and a backbone amide proton was inferred from the chemical shift change in the amide when the tyrosine was mutated to phenylalanine [45], but this was conclusive evidence given the mutation used. This approach, however, is not likely to be useful in most cases.

The hydrogen bond constraint is usually specified as a distance range between the hydrogen and acceptor atoms (1.8–2.1 Å). Linearity of the hydrogen bond may also be promoted by using a constraint between the acceptor and donor heavy atoms (2.7–3.0 Å for backbone N and O atoms). If structure refinement uses a force field that includes either an electrostatic term or an explicit hydrogen bond term (see Section 4.3 for more details), one tends either not to include hydrogen bond constraints at all, or just include those distance constraints between the hydrogen atom and acceptor atom (see also below).

Structures that are otherwise poorly defined may be significantly improved through the addition of hydrogen bond terms [30]. Therefore, hydrogen bond constraints are extremely valuable for calculating protein structures. For this same reason, evidence of regular secondary structure (from sequential NOEs) should not be reason enough to define hydrogen bond constraints based upon “normal” hydrogen bonding patterns. This has a high risk of biasing the outcome of any structure calculation.

Instead of adding explicit hydrogen bond constraints, it may be desirable instead to refine the protein structure using a force field that includes a good electrostatic or polarizability term. This approach would allow energetically favorable hydrogen bonds to form due to the influence of the force field. This can, in my experience, be somewhat

risky especially when the refinement is performed *in vacuo*. In this case it may be much more energetically favorable for charged groups to form hydrogen bonds or salt bridges at the expense of violating experimental distance constraints. However, analysis of constraint violations and comparison of observed hydrogen bonds with the observed slowly exchanging amide data is a reasonable approach for deriving high confidence and high quality structures.

### 3.5. Chemical shift constraints

Chemical shift is the most sensitive parameter to the structural environment, it is the easiest to measure with the highest precision and accuracy of all NMR parameters. Unfortunately, the origins of chemical shift are not fully understood, and currently inexact. Since such minor details have not prevented NMR users from using inexact and imprecise constraints derived from NOEs, hydrogen bonds or J-couplings, there is no reason to discount the utility of chemical shift in structure calculations!

Another problem that limits the use of chemical shift for structure calculations is that chemical shifts are not capable of yielding point-to-point constraints, such as those from interproton NOEs, J-couplings, or hydrogen bonds. Since chemical shifts are a product of the local environment, one can only confirm that an observed chemical shift is consistent with a calculated structure. Nonetheless, while this precludes a *predictive* use of chemical shift values, it does not eliminate their use when refining structures or determining their accuracy.

For proteins that have been assigned and have a high-resolution structure available, the root-mean square difference between the observed and calculated chemical shifts is about 0.3 ppm [46,47]. This precision is sufficient to confirm, at least, the orientation of strongly shifted resonances. In practice, chemical shift calculations have been useful for determining or confirming stereospecific assignments of leucine methyls [46] and glycine alpha [47] protons. Chemical shift calculations could, in principle, be useful for determining stereo-assignments whenever the chemical shift difference of the diastereotopic resonances differ significantly from random coil values. Given the importance of stereo-assignments for structure determination, chemical shift information is clearly valuable for constraints, although common use of "chemical shift constraints" in structure calculations has not yet occurred.

## 4. STRUCTURE CALCULATION STRATEGIES

NMR-based structure calculations essentially involves finding conformations of the molecule that are consistent with defined structural constraints. To perform this task, the researcher will face many decisions and options along the way. The four main decisions one is faced with are: (1) how starting structures are generated; (2) how the conformation is altered to optimally satisfy the constraints; (3) the description of the force field that alters the conformation; and (4) how the different calculation parameters (e.g., temperature, constraint weights, number of minimization cycles) are adjusted during the course of the refinement according to the calculation protocol. Some of the available options are described below.

### *4.1. Starting structures*

There is no means of taking a set of inexact and incomplete distance and torsion constraints and transmuting them into a protein conformation that satisfies all of the input constraints. Therefore an initial conformation — the starting structure — needs to be defined and then altered until the final conformation satisfies the input constraints as much as possible.

The choice of a starting structure can be crucial, depending on the means of refinement used and the number of experimental constraints available. For an absurd and extreme example, if the starting structure were the X-ray crystal structure, and if no NMR-derived constraints were used, and if the refinement means was molecular mechanics minimization, one would obtain a very nice looking final structure that did not differ from the “NMR constraints”. Yet this structure would not be meaningful, as the final conformation was due entirely to the choice of the starting structure.

The objectives in choosing a good starting structure is to first avoid any operator bias, second to assist in sampling the conformations that satisfy the constraints, and last to choose structures that can be refined fast enough to allow repeated calculations. Several calculations are required as NMR information is never sufficient to have the structure converge to a single conformation. Instead, one or more families of closely-related conformations will result from these calculations. This is the reason why it is essential to sample conformation space as much as possible.

#### 4.1.1. *Random conformations*

A common strategy is to use a randomly generated conformation as a starting point. For proteins, this may be accomplished by choosing random  $\phi$ ,  $\psi$ , and  $\chi$  torsion angles; this will be feasible only if any disulfide bonds that would be present in the folded conformation are left unformed.

Another option is the use of randomized coordinates for all atoms (the atoms are confined to a given volume). This approach requires significant time to recreate proper bond geometry, and there is no meaningful advantage to using randomized coordinates to using random torsion angles. Hence its utility is doubtful.

Using some form of random starting structure has the advantage that there is no user bias introduced into the starting conformation, and each random starting structure will assist in sampling conformation space, provided it is generated with a different random number generator seed. The disadvantages of using a random conformation is that one must use more computational time to refine the structure than would be needed if the starting conformation were closer to the final conformation. The other disadvantage is that it is possible to get caught in very high minima unless local structure is formed prior to global folding [48].

#### 4.1.2. *X-Ray or modeled structures*

The choice of a known X-ray crystal structure as the starting point for NMR-based protein structure determination is virtually unknown, although NMR-based nucleic acid structure determination studies have elected to use this strategy. It is one of dubious merit, chiefly in that conformational space searching is likely to be quite low, and in that it is difficult to assess the degree of structure “determination” when one starts with a structure that is already well-formed.

One of the earliest NMR-based structures determined, the lac repressor headpiece, used model building to generate a starting structure [49]. This was done by first fixing the backbone conformation of three alpha-helices (determined through sequential NOE information), and then by modifying the backbone torsion angles between the helices, until NOE constraints between the helices were (mostly) satisfied.

With today's commercially available protein modelling software, it would be feasible to generate modeled starting structures of larger and more topologically complex proteins (perhaps using databases of known protein conformations). The use of modeled structures as starting structures, however, is virtually nonexistent for two good reasons. First,



operator bias is quite likely unless the structures generated during modeling are driven by statistical methods, and not by expert or otherwise operator interaction. Second, these methods are simply too time consuming and difficult, given the alternatives.

#### 4.1.3. *Metric geometry (distance geometry)*

Another popular strategy is to use metric matrix (distance geometry, or DG) techniques to convert distance information into starting coordinates. Metric geometry techniques have been around since the 1930s [50] and are techniques where conformations that are described as distances between otherwise undefined points are converted into a coordinate system. In the original definition of metric geometry, points were explicitly declared to be undefined. Therefore I will limit the meaning of “distance geometry” to be only those manipulations that are done before conformations are generated or “embedded” into a Cartesian coordinate system.

Metric matrix methods involve creating a  $N \times N$  bounds matrix (B) for a molecule of  $N$  atoms, where  $B(i,j)$  is the lowest allowed distance, and  $B(j,i)$  is the longest allowed distance between atoms  $i$  and  $j$ . These are referred to as the lower and upper bounds, respectively. Bounds information is entered for known bond lengths and bond angles (described as distances), for all NOE distance constraints, and for all J-coupling derived torsion constraints, again described as distances. Next, the distances are used to further restrict the bounds using the triangle inequalities, in a process known as “bounds smoothing”:

$$L_{ik} \geq \max [(L_{ij} - U_{jk}), (L_{jk} - U_{ij})]$$

$$U_{ik} \leq U_{ij} + U_{jk}$$

where  $L_{ij}$  and  $U_{ij}$  refer to the lower and upper bounds between atoms  $i$  and  $j$ . Essentially, this process is checking to see that for all atom triples, the distances between them describe a triangle that is realizable in three-dimensional space.

Tetangle smoothing (where all possible tetrahedra are checked for bounds consistence) is also possible, but it is so computationally intensive that it is rarely done (unless one has access to much supercomputer time). Given that computing speeds are always increasing, it may be possible that tetangle smoothing will be feasible in the not too distant future for protein-sized problems.

Since bounds smoothing is the most computationally intensive step, and since a single smoothed bounds matrix can be used for repeated embedding steps, most distance geometry programs will allow the bounds matrix to be saved to a file after triangle inequality smoothing.

Next, trial distances are selected that lie between the upper and lower bounds, and it is in this step that different commercially available metric matrix programs diverge. In the simplest case, the trial distances are selected randomly between the upper and lower bounds. However, it is likely that the trial distances will no longer obey the triangle inequalities. One alternative is "metrization", where a trial distance is chosen for a bound, followed by triangle re-smoothing. This is repeated for some or all bounds. It is essential that the order of the bounds chosen for metrizing is random, else severe conformation sampling problems can occur, especially when the number of distance constraints is few.

The main effect of performing the additional smoothing steps is to yield a trial distance matrix that will yield better embedded structures (see below). However, it does so at a large cost in calculation time for the smoothing steps, especially for larger proteins.

The next step is to convert the trial distance matrix into a metric matrix, where the metric is the distance squared to the center of mass. This is calculated easily from the distance matrix. Last, the three or four largest eigenvalues are extracted from the metric matrix, and their corresponding eigenvectors are used as the basis set for converting all of the interatomic distances into coordinates.

What one gains from distance geometry is a starting structure that has much of the overall folding or topology present without any operator bias being introduced (this is because all distances have been used with equal importance in generating this structure). The embedded structure, however, can be very distorted, and must be optimized further using the methods described in the next section. The distortion is completely normal, as distance geometry/metric matrix methods will only create an exact structure if all distances are known exactly. This is never the case.

A last point concerns the effort spent on "metrizing" the bounds matrix, where trial distances are chosen so that they obey the triangle inequality [51,52]. This is a very time-consuming step. Furthermore, the problems associated with low conformation space sampling are most notable when the number of constraints are few (or zero), which is usually not the case. In my experience it is faster to use only random trial distances, and spend more time performing minimization and

molecular dynamics or simulated annealing. As will be discussed below, some form of dynamics is very desirable to enhance conformation space exploration.

## **4.2. Conformation drivers**

Starting structure conformations must be modified until the NMR experimental constraints are satisfied. One route is to minimize the potential energy, the second is to convert the potential energy into kinetic energy, and then to slowly lower the total energy of the system.

### *4.2.1. Molecular mechanics*

Molecular mechanics involves minimization of the potential energy of the protein. This is achieved through calculating the displacement of atomic coordinates so that the energy in the new conformation is lower than the energy in the previous one. The atomic displacements are determined from the potential energy gradients, although only the direction of displacement is determined; the displacement length depends on the implementation of the minimizer.

The disadvantage of using only minimization is that it samples conformational space poorly, and the final structures can easily be caught in high-energy minima. The final structure will also be very dependent upon the starting structure used. Minimization, however, is usually the final step in structure calculations that involve molecular dynamics, described next.

### *4.2.2. Molecular dynamics, simulated annealing, and Monte Carlo methods*

These related methods may differ in how the decision is made to displace the atomic coordinates, but otherwise they are very similar. They all have the advantage over minimization in that the potential energy can go “uphill” at times. Molecular dynamics is a method that is meant to simulate the real movement of molecules at a femtosecond time scale. Simulated annealing and Monte Carlo optimization are related in that both make no claims that the motions simulate reality, and in that Monte Carlo methods are often used in dynamical simulated annealing. The important differences between these methods is in how the target functions are defined (see Section 4.3).

Molecular dynamics involves solving Newton’s equations of motions for the atoms in the protein. The starting velocities can be chosen from

a Maxwellian distribution, be randomly chosen, or be derived from the potential energy forces on the molecule. Dynamics runs are often performed for 5–50 ps and at temperatures of 300–1200 K to ensure sampling of conformational space. Some expertise is required when setting up a dynamics run, lest too much kinetic energy be introduced suddenly, resulting in the molecule becoming severely distorted.

Equally important is that the dynamics run ends with a prolonged period of slow cooling, referred to as “annealing” or “dynamical simulated annealing” (named after the physical phenomenon of annealing that is observed when molten metal is cooled slowly). Cooling is the removal of kinetic energy from the system, and may be obtained mathematically either by coupling the protein system to a heat bath (in dynamics), or by use of the Metropolis algorithm (in simulated annealing). Failure to cool the system will likely result in the protein being trapped in a high energy minimum, whereas a good annealing protocol will yield low-energy structures.

Molecular dynamics and dynamical simulated annealing are well-accepted means for both obtaining low energy structures and for exploring conformation space. It is also common to use a distance geometry embedded structure that has been energy minimized as the starting structure for dynamics. Combined DG/MD calculations, however, can be quite time-consuming for larger proteins using the available and affordable processors of today.

#### 4.2.3. *Torsion drivers*

Molecular mechanics and molecular dynamics operate in Cartesian coordinate space; torsion drivers alter conformations by altering torsion angles about rotatable bonds. Torsion drivers usually are linked with minimizers, but can be combined with other methods, such as Monte Carlo methods. Torsion drivers may also be constrained grid search routines, but these are only useful when the number of rotatable bonds is few.

The torsion space approach most often used for protein NMR structures involves the use of a variable target function [53], and is found in a few commercially available products, such as the DIANA program [54]. The variable target function strategy is used in conjunction with a random conformation starting structure in order to produce well-formed structures. The variable target function strategy involves a penalty function that gradually seeks to satisfy constraints. At the start, the only constraints considered are those between residues close in the amino acid sequence (say, only those constraint violations be-

tween residues separated by no more than three amino acids). The length of these segments is slowly increased during the calculation, until all constraint violations between all residues are minimized.

Early implementations of torsion space routines suffered in that only a small percentage on the trial structures resulted in a final structure, especially for proteins with large amounts of beta sheet structure. This problem has been solved by first calculating initial structures, and then analyzing the local geometries in regions that satisfy the input constraints to determine the acceptable torsion ranges [55]. These are then used as additional, redundant constraints (since they are in fact determined by long-range constraints). The utility of these otherwise redundant constraints comes from the use of the variable target function: by satisfying these local constraints during the initial part of the structure folding, it ensures that the long-range constraints will be satisfied.

A major advantage of torsion drivers are that they are much faster than coordinate space methods, as fewer degrees of freedom are involved (about one-tenth of the degrees of freedom used in coordinate space). For example, bond angles and bond lengths are kept at the initial values. The increase in speed is especially noticeable for larger proteins. Another advantage is that, as starting structures are random conformations, conformational space is presumably well sampled.

### 4.3. Force field choices

Minimizing the energy of a structure depends, of course, on how the energy is calculated. The potential energy of a system is often represented as the sum of two or more individual energy terms, namely, the sum of the bonded energy terms, the non-bonded energy terms, and the constraint terms (which may be between non-bonded or bonded atoms). A common model potential energy function is

$$V_{\text{tot}} = \sum_{\text{bonds}} \frac{1}{2} K_b (b - b_0)^2 + \sum_{\text{angles}} \frac{1}{2} K_\phi (\Phi - \Phi_0)^2 + \sum_{\text{torsions}} K_\theta [1 + \cos (n\theta - \delta)]$$

$$+ \sum_{\text{nonbonded}} \left[ \frac{A}{r^{12}} - \frac{C}{r^6} + \frac{q_1 q_2}{kDr} \right] + \sum_{\text{constraints}} K_d (d - d_0)^2$$

This treats the bond lengths and bond angle with Hooke's law, where each bond length or bond angle has a characteristic force constant  $K$  and

equilibrium value  $b_0$  or  $\phi_0$ . The third term accounts for the energy about rotatable bonds, and the fourth term accounts for both Lennard–Jones attractive and repulsive forces and electrostatic forces between atoms that are not directly bonded. The last term is a harmonic constraint term; other forms, such as a flat-bottomed harmonic well, are often used.

Force fields may be highly parameterized descriptions such as CHARMM [56], AMBER [57], and Gromos [58], among many others. The authors of these force fields have determined values for the various force constants and variables used in models that describe the potential energy of the system (usually similar, but not identical to the one above). These force fields attempt to be a complete description of the forces on the molecule such that the force field can predict spectroscopic or other physical properties. The molecule in question may be limited to being a protein, and the same force field that predicts protein properties well may be of limited use to carbohydrates.

Force fields may also be more simple descriptions that do not attempt such difficult goals as predicting protein structures and dynamical properties. For example, the energy change for deviations from standard bond geometries may be calculated with the harmonic potential shown above, but where all bonds have the same force constant. These simpler force fields may not predict physical properties with great accuracy, but they may be more useful for a wide variety of molecules and may find local minima that are structurally reasonable and within the range of conformations that are in agreement with the NMR constraints.

#### *4.3.1. Using geometric and constraint terms only*

The simplest useful force field is one that would maintain ideal bond lengths and bond angles, keep the closest separation between non-bonded atoms to the sum of the hard-spheres van der Waals radii, and enforce all experimental constraint terms. This formulation will emphasize the role of the experimental constraints in determining the final structure. This approach is also very widely used, for both simulated annealing procedures [6] and torsion space minimizers [53].

This approach makes a lot of sense for use in NMR-derived structure determination. For one, it does not make sense to include, much less emphasize, terms such as electrostatics and van der Waals attractive forces in a structure calculation until the structure being calculated closely resembles the true structure. If such terms were used in the earlier stages of structure calculations, it would not only slow the

calculation due to the increased number of energy terms that must be evaluated, it will also likely slow the calculation due to premature condensation of the structure.

#### 4.3.2. Chiral constraints

If the possibility of chiral inversion is present, as it is in DG or after four-dimensional annealing, then it is essential that the proper stereochemistry about chiral centers is maintained. This is also true for prochiral centers when stereospecific assignments are available for the diastereotopic substituents. Chiral constraints may be imposed in two ways: by defining improper torsion angles, or with a triple scalar product (also known as a signed volume).

#### 4.3.3. Including non-bonded terms

These terms include the Lennard–Jones attractive and repulsive forces, electrostatic forces, and a hydrogen bond term (which may be included in the electrostatic forces). As mentioned above, the use of these forces, except for the repulsive van der Waals contribution, on a random starting conformation does not seem sensible. The use on a structure that is highly refined and has an abundance of satisfied experimental constraints is reasonable, in that structures with lower energies, or that “make good chemical sense” may be determined. Furthermore, any contacts that are stabilized by such forces, and not by experimental constraints, may be tested through back-calculation of the experimental data. One must be careful to identify structures that satisfy the artificial force-field terms better than the experimental constraints [59].

#### 4.3.4. Variations in NOE constraint terms

Several different treatments of NOE distance constraints have been reported. The most common treatment has been to use lower and upper bounds for the NOE constraint, and to treat violations with a simple flat-bottomed harmonic potential:

$$V_{\text{noe}} = \begin{cases} \sum_{\text{constraints}} K_{\text{noe}}(d_{ij} - d_{l0})^2 & \text{For } d_{ij} < d_{l0}; \\ \sum_{\text{constraints}} K_{\text{noe}}(d_{ij} - d_{u0})^2 & \text{For } d_{ij} > d_{u0}; \\ 0 & \text{For } d_{l0} \leq d_{ij} \leq d_{u0} \end{cases}$$

A problem with this approach is that large upper bound errors result in huge forces being generated between the distant atoms. If these are used in, for example, a molecular dynamics program, the resulting velocities may cause to protein structure to have severe problems unless the velocities are rescaled. One might also treat this problem by linearizing the force potential above a given threshold value.

Another option, used when NOEs are converted to an equilibrium distance, simulates the sixth-order relationship between NOE intensity and distance [26].

$$V_{\text{noe}} = \sum_{\text{constraints}} K_{\text{noe}} \left[ (r_{ij}^{-3} - r_{0ij}^{-3})^3 - r_{0ij}^{-6} \right]$$

This form also has the benefit of creating small forces when the upper bound is strongly violated (as opposed to the previous function, where the forces increase strongly with increasing violations). The fact that it uses an equilibrium distance instead of distance bounds means it will be sensitive to errors in the NOE distance estimation. Nonetheless, if distances are refined using back calculation (see Section 5.3), this may prove to be an attractive alternative to setting distance ranges.

Both of these options use the approach that the “best” structure has all distance constraints in the structure simultaneously satisfied at every step in the minimization or dynamics trajectory. Since proteins are highly flexible molecules, it is more reasonable that the distance constraints should be treated so that a weighted average over time is satisfied:

$$d_{\text{noe}} \approx \langle (d(t))^{-6} \rangle^{-1/6}$$

This “time-averaged” approach appears to be superior for defining the structure and flexibility of the molecules [60–62].

Another method of growing interest is to directly refine against the NOE, which is calculated using relaxation matrix techniques (see also Section 5.3.1). One proposed potential [63,64] is

$$V_{\text{noe}} = \left[ \sum_{\tau_m} \sum_i W_{\tau_m, i} \right]^{-1} \sum_{\tau_m} \sum_i W_{\tau_m, i} \left( A_{\tau_m, i}^{\text{cal}} - A_{\tau_m, i}^{\text{obs}} \right)^2$$

where



$$W_{\tau_m,i} = \left( N + \varepsilon A_{\tau_m,i}^{\text{obs}} \right)^{-2} \quad \text{and} \quad A(\tau_m) = \exp(-\mathbf{R}\tau_m)$$

and where  $N$  and  $\varepsilon$  represent the experimental noise level and relative error in the experimental NOEs, respectively. The sum is for all peaks  $i$  at different mixing times  $\tau_m$  and  $A$  are the NOE intensities calculated from the relaxation matrix  $\mathbf{R}$ . This potential is used to calculate a gradient suitable for direct refinement of atomic positions. A related approach for use in torsion space optimization routines has also been proposed [65]. A problem with this potential is the large forces calculated when the initial structures has underestimated distances. This may also be treated by linearizing the NOE potential above some threshold value, as it could be in other potential forms.

The advantage of direct methods is that they assure a best fit between the model and the experimental data; essentially this combines NOESY back calculation directly into the structure optimization stage. This approach is analogous to the constrained least squares refinement procedures in crystallography that utilize a potential based upon structure factor differences:

$$V(\mathbf{F}) = \sum_h \left( \frac{1}{\sigma_F} \right) (|\mathbf{F}_{\text{obs}}| - k |\mathbf{F}_{\text{calc}}|)^2$$

which is a long-accepted practice.

Despite the apparent attractiveness of direct refinement methods, there are some serious drawbacks. One trivial drawback is that they are computationally very expensive. More seriously, they rely upon accurate NOE measurements, which are rarely obtained in practice, and require an unknown scaling term between the theoretical and experimental NOE intensities. Furthermore, the best-fit between model and experiment does not mean that the model is accurate, as the experimental data is likely to contain many inaccuracies. The improvement in the quality of the structures is usually less than the accuracy present, and given the cpu-time expense of direct refinement methods, they are of questionable utility.

#### **4.4. Optimization protocols**

Depending upon the nature of the force field and conformation driver used, there will be different parameters that the user can adjust during the course of the structure optimization. These may be obvious ones,

such as the total number of minimization steps to use, or the temperature to use during the dynamics run, or they could be more obscure ones such as the type of minimizer used or the dielectric constant used in calculating the electrostatic forces. How these parameters are adjusted is often the most critical factor in making the calculation a success.

#### *4.4.1. Beginning the optimization*

The major point of concern in the initial stages of refinement is that the experimental constraints not be applied too quickly. If all constraints are brought into play early on, it is possible to inadvertently “trap” the structure in an undesirable conformation [48], or to limit conformation space sampling. The exact remedy to use depends upon the type of starting structure used.

If the starting structure is a DG embedded structure, then the simplest and most reasonable way to limit the effect of the experimental constraints is by reducing the constraint weight to a low or zero value. DG embedded structures have extremely distorted bond lengths and angles; initial refinement is aimed at cleaning up the local geometries before trying to improve the global folding. Another initial concern is if the embedded structure has a mirror image topology relative to the correct structure (which would be known after the first few trial structures are calculated). If the embed is a mirror image, it is best to invert the embed prior to any other optimization step [59].

If the starting structure is a random conformation, then the problem to avoid is the premature condensation of the structure due to NOE constraints between atoms that are far apart in the starting structure. The preferred solution to this problem is to selectively weight constraints given the separation in the protein sequence of the two residues involved in the NOE. Typically, NOEs between residues farther apart than five positions are not included in the initial refinement, allowing the local structure to form first. This is similar to the procedure one would follow if folding up a plastic protein model [53].

If molecular dynamics are to be used for conformational searching, it is recommended that the experimental constraints have a low constraint weighting, say in the order of 0–5 kcal/Å, for the first few picoseconds of the dynamics run. This will help keep the constraints from damping the initial dynamics trajectories, allowing more conformation space to be sampled. In addition, warming the system up over 1–2 ps is advisable.

#### 4.4.2. Optimization: getting the constraints satisfied

The bulk of the structure optimization is concerned with changing the initial starting conformation to a new conformation that satisfies the experimental constraints as much as possible, without distorting the rest of the structure. During this time, the number of constraints used is increased (if the long-range constraints were ignored initially), and the weight on the constraints relative to other forces also is increased. Both should be done slowly, but not so slowly as to take excessive time in doing the calculation. The rate used may be determined by trial and error, or by prudent choice. For example, the long-range distance constraints should not be used until the trial conformation has reached a point where helices and turns are formed, and the initial condensation of the structure reached. Also, if the constraint weight is so low that the forces generated by the constraints are not sufficient to converge to the final conformation, then the weights need to be increased.

Other means may be required to overcome folding barriers that occur when parts of the conformation are trapped in high energy local minima, which can occur even if care was taken not to emphasize long-range NOE information earlier. Two methods have proved to be useful to untangle knots in the local geometry. The first is to reduce the van der Waals radii (used for determining closest approach). The radii can even be reduced to zero, allowing atoms to pass between each other, as the energy barrier between the improper and proper folded states is reduced.

High energy barriers can also be reduced through the use of four-dimensional dynamics or minimization (the fourth dimension is simply another coordinate, perhaps originating from a fourth eigenvalue during distance geometry embedding). If no weight is placed on the magnitude of the fourth dimensional coordinate, then atoms may also freely pass between each other. This not only reduces the effective van der Waals radii, but also allows other energy barriers to be overcome. Molecules in this four-dimensional space can be returned to three-dimensional coordinates by slowly reducing the penetration into the fourth dimension, or by obtaining a three-dimensional projection at any point (by zeroing the fourth dimensional coordinate).

If distance geometry was used to generate a starting structure, or if four-dimensional dynamics have been used, then one must also ensure that the proper configurations are obtained for all chiral centers, and for prochiral centers when stereospecific assignments are known. This

should be done early in the calculations. If proper chirality is not enforced automatically by the force field used, then it becomes necessary for all chiral configurations to be manually checked and inverted, if necessary.

#### *4.4.3. Obtaining conformational sampling*

Since NMR data does not lead to a unique structure being determined, it is important to determine the range of conformations that are consistent with the experimental data. This is dependent in part upon the steps that have been taken prior to optimizing the structures, namely in the choice of starting structure. It is also possible to enhance conformation sampling using dynamics (or annealing) procedures for optimization. The parameters that regulate conformation sampling are the length of the dynamics run, the temperature used (the higher the temperature, the greater the velocities, and the increased likelihood of enhanced conformation sampling), and the force constants used for the constraints.

High force constants for the constraints will act to dampen the motions of the molecule. They may also restrict the structure to a particular conformation that is not necessarily the only satisfactory conformation. For example, conversion between different loop conformations in a protein may only be possible if the protein can “breathe” somewhat. If this meant that several constraints be violated temporarily, then high constraint force constants would prevent the protein from breathing, and hence the different loop conformations from interconversion. For this reason, constraint force constants should be slowly increased during the course of the optimization.

If the method of floating chirality mentioned in Section 2.3.3 is used, then it is important that a good statistical sampling of both configurations is obtained. This can be only obtained if the initial starting structures have random starting configurations, or if interconversion between diastereomers during optimization is allowed. Interconversion usually does *not* occur when normal simulated annealing, dynamics, or torsion space refinement is performed. Interconversion can be achieved either through direct manipulation (swapping the coordinates of prochiral hydrogens, for example), or if four-dimensional optimization is used. Chiral conversions in four dimensions become a simple rotation, where bonding geometry should not restrict conversion between pro-R and pro-S conformers.

#### 4.4.4. Final stages of optimization

If dynamics or annealing have been used, then the temperature should be lowered slowly, and followed by minimization. Usually the final annealing steps are the most critical for obtaining high-quality structures. The annealing must not be too rapid, or else “quenching” of higher-energy structures will result. Minimization itself is completed when the gradients are too small to continue, or some number of iterations is exceeded.

Optimization should yield low-energy structures that are a good fit to the constraints. Therefore all constraints should be used at their full force constants. A typical NOE force constant would be on the order of 30 kcal/Å, or roughly 15% of the average bond stretching energy. Using lower force constants may result in structures where the constraint violations exceed desired levels. Usually, constraint violations greater than 0.3 Å are not desired, and structures with such high violations may be either discarded or optimized further. This is discussed in more detail in the following sections. A possible protocol is shown in Fig. 2.

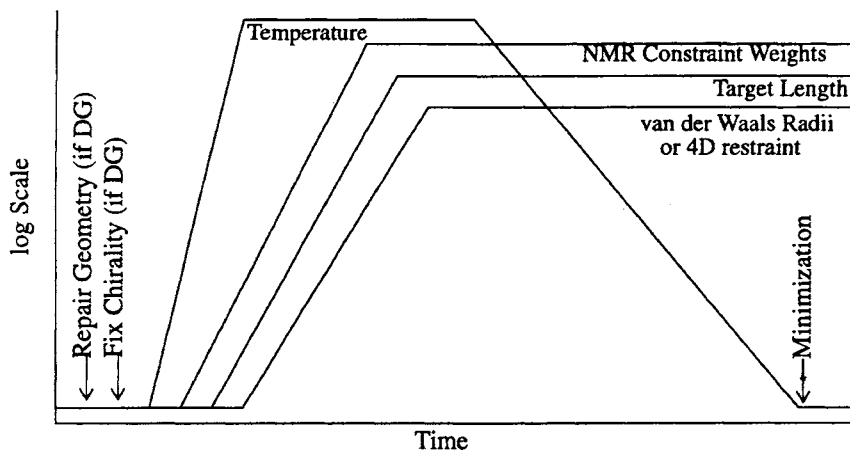


Fig. 2. Stylized schematic representation of an optimization protocol. Lines represent relative paths for increasing NMR constraint weights (NOE distances and torsion angles, etc.), the target length in a variable target function, the van der Waals radii or fourth dimension penalty, and the dynamics/annealing temperature (if applicable). See text for more details.

## 5. ANALYSIS OF THE CALCULATED STRUCTURES

No structure determination should be complete without a critical assessment of the validity of the final structure(s). This is important, as coordinate files can be used in very misleading ways. For example, they can be used to create absolutely stunning color pictures using modern graphics workstations. These photos can be absolutely meaningless if their creator has not been critical in evaluating the correctness and meaning of the structures.

### ***5.1. What is accuracy? What is a structure?***

In my opinion, it is how we represent our data as structures, and how we report our errors (or goodness) that is most indicative that NMR structure determination is still in its infancy.

We often strive to describe the accuracy of calculated structures in a variety of ways. A common, but completely erroneous way is to point to an RMSD value for several independently calculated conformers. This is wrong for several reasons, the first is that it has nothing whatsoever to do with accuracy. RMSD values are more of a measure of precision, but even this is very risky as structures can converge even when precision is poor. Convergence can be due to either convergent data, or a minimization (structure refinement) routine that will always end up in a similar local minimum. I have seen excellent convergence when using DIANA, even though several constraints are severely violated (due to having inconsistent constraints in the molecule). Therefore excellent convergence of the conformations can result due to a deep local minimum encountered in the polydimensional topological representation of the data. To say that this convergence is a measure of either precision or accuracy is quite wrong, as the accuracy is clearly bad, and the precision can be dictated (at least in smaller systems) by only a few constraints.

Another problem with using RMSD values is that they usually are derived in a subjective manner. Typically the fits performed do not use all residues in the molecule, as some regions may be locally or globally disorganized; the user decides which residues to include for determining the global RMSD values. The user will select residues based upon perhaps two criteria: (1) which residues *appear* to be poorly fit when a global fit using all residues is applied; and (2) which residues, when excluded, make the resulting RMSD value closer to what the user wants. This second criterion is one that makes RMSD values clearly

suspect for use as reporting structure “goodness”, but even the first criterion is not rigorous.

Measurements of accuracy and precision also assume the misleading paradigm of representing NMR data as a single structure. Molecular motions on the NMR time scale will result in the protein sampling a conformational space that cannot easily be represented or understood by the user. So instead we use the attractive alternate of representing our data as a single (perhaps averaged) structure. We do this because, like the person looking for a lost object in the wrong place, we find the light is better when we look at a single structure.

## **5.2. Residual constraint violations**

One of the more compelling indicators of structure validity is the number and magnitude of constraint violations present in the structure. If there are many violations and the magnitude of the violations is high, this indicates that either the structures are not completely optimized, or that there are serious problems with the constraint set, and quite possibly errors in the resonance assignments. These can be rather common problems that occur when using manual bookkeeping of the assignments and constraints.

The number of constraint violations is also only meaningful if an acceptable number of constraints is present. Typically 15 or more constraints per residue should be present for a structure to be well-determined. This number is higher for systems studied by three- and four-dimensional heteronuclear methods, as many more aliphatic NOEs can be identified.

Acceptable constraint magnitudes also depend upon how the constraints were defined. For example, if very wide constraints were used (say 1.8 to 5 Å for every NOE), then virtually any violation would be intolerable. If, however, very narrow constraints were used (say 1.8 to 2.1 Å for a strong NOE), then violations of 0.2–0.3 Å may be more acceptable, and the user should reevaluate whether such a narrow constraint was justified, or refine the distance estimates using back calculation procedures.

One must also question whether it is reasonable, let alone possible to satisfy all of the distance constraints, as almost certainly the molecule of study is undergoing motions that lead to time-averaged NOEs and J-couplings. These may not be consistent with any single structure.

The constraint violations encountered also should be interpreted with respect to the stage of the structure determination. Initial structures

using fewer constraints may be determined that have some substantial violations, but structures that are being considered as final and refined should not. One significant part of structure determination is the determination of accurate and precise constraints; this often requires iterations between preliminary structure calculations and re-interpretation of the data sets. This is described in the next section.

### ***5.3. Back calculating spectra from structures***

Satisfying constraints does not necessarily mean that the structure calculated is accurate. One also needs to show that the structures are in agreement with the original data, since the constraints are a subset of the information present in the original data. Specifically, it needs to be demonstrated that:

1. No appreciable NOEs are predicted to appear where none are observed.
2. NOEs observed but not used as constraints are predicted to appear.
3. The relative intensities of predicted NOEs agree with the experimental data.
4. Predicted J-couplings agree with the experimental data.

One might consider testing if the resonances that are strongly shifted from their random coil values have calculated shifts that are in agreement with the observed shifts [47]. While this approach is starting to be developed, it is still too early to tell how well it will work and when it will be generally available.

#### ***5.3.1. NOESY Calculations***

Calculating a NOESY spectrum from a model structure is the usual NMR “back calculation”. This task involves solving Solomon’s equations describing cross relaxation for the entire molecule [67]. It is highly desirable to perform this type of back calculation, as it will take into account spin diffusion (which otherwise hinders correct interpretation of the distances). Two approaches are possible:

1. Numerical integration, using a sufficiently small time step is needed. Used in conjunction with a cutoff for considering dipolar interactions, the efficiency is order  $O(N^2)$  for a system of  $N$  resonances.
2. Matrix approaches that transform a matrix of intensities into a matrix of cross relaxation rates using diagonalization. These have an efficiency of order  $O(N^3)$ .



At first glance it would appear that numerical integration is superior [68]. This is a little deceiving, as it does not take into account the number of time steps needed for integration. It is faster to use matrix methods for calculating cross relaxation out to long mixing times (such as 1.0 s). Analysis of cross relaxation at long mixing times may not be important, as the effects of spin diffusion and the influence of secondary structure can be observed at relatively short mixing times [69].

Matrix methods require intensities for all NOESY peaks, including the diagonal peaks, in order for the calculation to be carried out. This is not possible for protein spectra, where the overlap is too great to extract all peak intensities. Nonetheless, it is possible to use a hybrid matrix that contains the available experimental intensities, and where missing intensities are computed *a priori* using a model structure. Iteration can then be performed between the calculated and experimental intensities until convergence is reached (according to some preset criteria).

This approach is used in the IRMA [70] and MARDIGRAS [71] programs, although how the iteration is done varies between them. IRMA iterates using the relaxation matrix approach to provide refined distance constraints that are then used in restrained molecular dynamics and/or distance geometry to provide a new model structure. MARDIGRAS iterates only the relaxation matrix diagonalization, where the experimental intensities are combined with the new theoretical intensities, until convergence is reached. Both IRMA and MARDIGRAS are reported to converge in only a few cycles, and provide results that are independent of the starting structure used.

Back calculation methods can (and should) take into account various forms of molecular motion, for example aromatic ring flips, or methyl rotation. Fully accounting for all relaxation pathways (including local motion, chemical exchange and taking into account the leakage rate) is the main shortcoming of the current implementations of NOESY back calculation schemes for biopolymers. Too much reliance can be placed on relaxation models that contain flaws. If back calculation is used where input distances are obtained from cross peaks that are involved in some form of chemical exchange, then the back calculation will yield NOESY spectra that are self-consistent even though the calculated structures are wrong [72].

### 5.3.2. *J-Coupling calculations*

Although *J*-coupling derived torsion constraints are very important and widely used in NMR structure determination, no effort is usually

made for back calculating correlation spectra from structures. Instead, reliance is placed on correctly identifying all possible coupling values from the data, and having the corresponding constraints be satisfied during structure optimization.

The drawback to this is that any couplings that have not been included in the constraint list may not be adequately predicted by the structure. Another case that may warrant checking is if the experimental couplings are motional averaged, yet the torsions in the calculated structures are always near some given value. This may indicate an overemphasis on other constraints that leads to “determination” of some local structure that is disordered in solution.

#### **5.4. Structure convergence**

Since the NMR data are not sufficient to determine a single conformation exactly, we try to test how well converged the resulting structures are using RMSD comparisons. As described in Section 5.1, this is not necessarily a valid analysis, but it is one of the few analyses readily performed, and so it is. Most often this is done by finding the best fit superposition between multiple structures, and reporting the RMSD values between them. Well-converged structures will have RMSD values for backbone atoms of less than 1 Å, and for all atoms it should be less than 2 Å. Some structures reported have RMSD values that are notably lower than these values, but given the problems in assessing how informative these RMSD values are, the values mentioned are offered as general rules of thumb.

If regions of the protein are disordered, these regions are usually excluded from consideration in both calculating the best fit and therefore in the RMSDs. The user will determine which regions are disordered by first superimposing structures using all atoms. Visual inspection usually will show which regions poorly defined. In extreme cases there will be not global structure calculated, while there is appreciable local structure; for example, two well-formed helices connected by a flexible linker region will not yield good global overlaps, but the individual helices overlay well.

To remove operator bias in determining which regions to include in the superpositions, it is possible to use the statistics obtained in superpositioned structures as a basis for discriminating between well-formed and disordered regions. Such an approach has been previously reported [66]. The only note of caution in this approach is that the user still determines the statistical thresholds for specifying what constitutes a

disordered region. This parameter can be adjusted to tailor the result to the user's preference, and therefore could be used in misleading ways.

### 5.5. Standards for structure determination

The previous points are raised to demonstrate the need for a clear, unbiased and informative means for assessing the quality and accuracy of structures derived from NMR data. Unfortunately, it is not clear whether this is possible with the kinds of information that NMR yields, and in the way that structures are determined.

One often sees reports proposing an "NMR R-value", similar to the R-value used routinely in reports of X-ray crystallographic structures. Claims that calculated structures were equivalent to an  $n$  Å resolution crystal structure have also been heard. Since NMR and X-ray crystallography are the only methods for determining protein structures at atomic resolution, and since X-ray has been in this business for a much longer time, it is understandable that these comparisons to crystallography are made. However, the nature of NMR and crystal samples, data, and structure calculation methods vary so much that perhaps these comparisons are inappropriate.

Ångstroms resolution in a crystal structure is inversely proportional to the size of the sphere of reflection used for collecting data, and therefore the number of reflections used in calculating the structure. No clear correlation exists for claiming an "Ångstroms resolution" for NMR data, although clearly more NMR constraints are required for increasing the "resolution" of the structure. However, the relationship ends there. For example, a twofold increase in resolution in crystallography requires the collection of *eight* times as many reflections. The difference that a single NOE or torsion constraint can make on the structure of an amino acid can be quite appreciable.

The dissimilarity between a crystallographic R-value and an NMR "R-value" is also notable. The R factor in crystallography is usually expressed as:

$$R(F) = \frac{\sum \|F_{\text{cal}} - |F_{\text{obs}}|\|}{\sum |F_{\text{obs}}|}$$

R-values can be misleading. One neglected aspect of this R value is that it is sensitive to the level of substitution in the particular derivative.

Metal-containing proteins will also give rise to R values that are deceptive. A similar R-value proposed for NMR is:

$$R(\text{noe}) = \frac{\sum_{\tau_m} \sum_i |A_{\tau_m,i}^{\text{calc}} - A_{\tau_m,i}^{\text{obs}}|}{\sum_{\tau_m} \sum_i A_{\tau_m,i}^{\text{obs}}}$$

where the meanings of the variables are the same as those described in Section 4.3.4. This NMR R-value is also deceptive, in that it will be dominated by short distances. In addition, the maximum value of this function is infinity, hence it is not possible to compare it to the crystallographic R-value, which has a maximum value of 0.83.

One possibility is to incorporate the sixth-power dependence of the NOE into the above NMR R-value:

$$R(\text{noe}) = \frac{\sum_{\tau_m} \sum_i |(A_{\tau_m,i}^{\text{calc}})^{1/6} - (A_{\tau_m,i}^{\text{obs}})^{1/6}|}{\sum_{\tau_m} \sum_i (A_{\tau_m,i}^{\text{obs}})^{1/6}}$$

Many other similar forms are possible [73], of which some appear to have reasonable interpretive use. These may also be used to create analytical functions suitable for incorporation directly into structure calculations, as was mentioned in Section 4.3.4.

It has been noted that all of these R-values have the hidden bias in that their values depend upon whether an experimental value is more intense than the calculated value, or *vice versa* [74]. The example given is that if the observed and calculated values have relative intensities of 1 and 3, then the R-value would be either 2/3 or 2, and it is not obvious how to interpret the result, or more importantly, how to use it in refinement. The alternative proposed is to use a root mean square approach. This would modify the previous R-values to be:

$$\text{RMS}(\text{noe}) = \left\{ \frac{\sum_{\tau_m} \sum_i [(A_{\tau_m,i}^{\text{calc}}) - (A_{\tau_m,i}^{\text{obs}})]^2}{\sum_{\tau_m} \sum_i (A_{\tau_m,i}^{\text{obs}})^2 + \sum_{\tau_m} \sum_i (A_{\tau_m,i}^{\text{calc}})^2} \right\}^{1/2}$$

### **5.6. Accuracy of NOE peak volume measurements**

If back calculation methods and the direct refinement methods (described above) are to use NOE intensities in their calculations, then it becomes important that these intensities be measured accurately. Such accuracy is not required if one is only going to treat NOEs qualitatively when defining constraints, and one is not using back calculation to check the accuracy of the structures.

If the back calculation is used in conjunction with lineshape models to generate simulated NOESY spectra, then it may be possible to compare the experimental and theoretical intensities on a point-by-point basis. In this case, one may start with relatively imprecise distance constraints to generate the initial model, and then refine based on the actual data values. Since this would require correct lineshapes and chemical shifts for all resonances, this option does not seem currently feasible. Hence it necessary to weight all NOE contributions with respect to the estimated error in the NOE measurement.

## **6. FUTURE PROSPECTS**

In the introduction to this chapter, I stated that the field of NMR-based protein structure determination is still in its infancy. From the previous discussions, I hope that this statement is both understood and accepted by the reader. I also made this comment in the light of recently becoming a first-time father. Infants are incredibly well developed and complex beings, so it should not be an insult to anyone to say that this field is in its infancy. It is without any doubt, however, a field that will continue to rapidly grow and develop (and like an infant, have an occasional dirty diaper). In this section some of the areas of special promise are discussed. (Note added in proof: since originally writing this chapter, my son has moved into his “terrible twos”. More than ever the analogy to infancy seems appropriate, as just when we think we might be able to “control” the problem, we find the problem has a mind of its own. But then, to quote Piet Hein, problems worthy of attack prove their worth by hitting back.)

### **6.1. Automation of resonance assignments**

Resonance assignments are the first, the most important, and perhaps most tedious step in the path to determining a structure from NMR spectra. Currently, assignments are obtained mostly through

manual inspection of the NMR spectra, although several labs have reported automated or computer assisted resonance assignments [75–82]. It is expected that this trend towards automation will continue, and that the number of resonances determined automatically will grow. The benefits of automation are not only the increase in speed, but also the increase in reliability when efficient bookkeeping schemes are combined with intelligent automation programs.

#### *6.1.1. Routines are needed to automatically create accurate, reliable peak lists*

Resonance assignment automation involves some form of pattern recognition, where the data are peak lists obtained from the spectra, and the patterns searched for correspond to through-bond and/or through-space connectivities. Since the basic information used is peak lists, and not the original data, it is imperative that the peak lists be of very high quality. Peak picking must be able to treat noisy or artifact-ridden spectra, and overlapping peaks must also be treated correctly. The adage of “garbage in, garbage out” is often encountered when trying to automate assignments with poor quality peak lists.

A high-quality peak list is one that is free from artifactual peaks and has very precise peak centers defined (within 0.005 ppm for proton shifts, 0.05 ppm for carbon and nitrogen shifts).

Neural networks or other non-linear pattern recognition methods are a possible approach to this problem. These appear to be a good choice for further research, where neural networks of sufficient depth and complexity are trained on data sets that have been pre-peak picked manually. One can also envision training different networks to recognize peaks in different types of spectra, such as DQF-COSY, NOESY, TOCSY, etc., such that each network can be adequately trained to handle a specific data type. Fortunately, there exists considerable spectra suitable for training such networks.

Simulated annealing or related genetic algorithms for extracting high quality peak lists are also methods that one might predict would be very useful when applied to the problem of getting peak lists, especially since these methods might be able to use the resulting assignments as the “target function” for the peak list extraction. That is to say, since the assignments are dependent upon good peak lists, then a “gradient” for driving peak list extraction could be the quality of the resulting assignments.

Currently the best approach appears to be combining the pattern

recognition ability of a trained user with a software package that allows for easy, interactive (and often manual) preparation of a high quality peak list.

#### *6.1.2. Efficient means of storing peak information and derived knowledge are needed*

For a typical protein NMR structure problem, one will be dealing with several multidimensional spectra, each containing hundreds to thousands of cross peaks, which are to be assigned to hundreds to thousands of nuclei, and be used in generating a like number of structural constraints. This is a well-recognized bookkeeping problem.

Some of my own research has been directed towards this end. Clearly this problem is best approached using software to perform the bookkeeping; manual forms (and nonelectronic forms) are too prone to transcription errors, or other common human mistakes. The form I favor has been implemented in a commercially available product under development in my group. Bookkeeping is handled by a spreadsheet, where each row in the spreadsheet is a peak, and columns are peak attributes. We have created column types to handle every type of data known, plus columns that are evaluated at run time. The spreadsheet can be dynamically linked to spectral data or molecules, as well, such that relationships between spectra, peak list, and structure can be queried. How this spreadsheet may be best used for automation and bookkeeping is an ongoing focus of the group. For example, we are looking at ways to use interaction between plotted spectra and spreadsheets in order to assess peak list quality, and accuracy of automatically-derived assignments.

#### *6.1.3. Pattern recognition algorithms are needed*

Much of resonance assignments can be described as finding peaks that line up in a common dimension. This corresponds to a simple algorithm to implement in a computer code. Basically, if the difference in chemical shift between the two peaks ( $i,j$ ) in a given dimension  $n$  is less than a tolerance  $\epsilon$ , then the peaks *may* be related. Except in the simplest spectra, the next step is to determine whether or not the alignment is due to actual coupling between the resonances involved, or is a consequence of spectral overlap. This decision will be based upon additional information, such as do the two peaks have the same lineshapes in the common dimension, or whether the two peaks also are aligned in spectra acquired under different sample conditions. It can

also be done by correlating information between several different experiments. For example, the reliability of a connection between an  $\text{NH}-\text{C}^\alpha\text{H}$  cross peak and a  $\text{C}^\alpha\text{H}-\text{C}^\beta\text{H}$  cross peak in a DQF-COSY spectrum is very high if the corresponding  $\text{NH}-\text{C}^\beta\text{H}$  cross peak is found in a TOCSY or RELAY spectrum [78]. Additional information for determining if two cross peaks are related is knowing whether or not you expect them to be related. This type of information is often used when making manual assignments — for example, one distinguishes spectral overlap when there are too many peaks that line up, and so only the “best” possibilities are used. This pattern recognition step could be encoded so that an automatic assignment routine would be able to use the molecular knowledge to help determine the assignments.

Further progress in making the resonance assignments is in the use of heteronuclear three- and four-dimensional experiments to obtain couplings throughout the protein backbone [4], as described in Section 2.2. I have found these experiments very easy to automate when using the spreadsheet approach developed in my group in conjunction with a flexible macro language. We are also investigating the application of graph theory algorithms to represent heteronuclear spin systems, mapping them to the known molecular structures using graph/subgraph matching [79].

#### *6.1.4. Automation algorithms need to be easily modified or developed, given the rate of pulse sequence propagation*

Automation algorithms are written such that they interpret specific experiments. For example, a program may compare DQF-COSY, TOCSY and NOESY data to derive sequential assignments. The same program most likely would have no means of dealing with the 3D HNCA, HNCB, etc. experiments described in Section 2.2. Since it is not uncommon for new experiments to be developed quickly, and for these experiments to rapidly propagate to other laboratories, it seems clear that automation routines must be easily adaptable for new experiment types.

The best way to approach this problem, in my opinion, is to have a software toolkit that has bookkeeping and logic tools available in a macro language. This will allow researchers to “program” NMR/modeling software by assembling different basic tests (such as do the peaks line up, or are two atoms close to another in the molecule) into algorithms that duplicate the steps done by the expert when manually analyzing information. For example, programs that Luciano Mueller and I developed in VMS Fortran [77,78] would require considerable



modification to handle 3D data, and it would have to be done in Fortran. Alternatively, the same capabilities are easily obtained and modified by a researcher using the program commands and macro language available in the SYBYL software, which my group has been involved in developing.

We are also investigating the use of a simple line notation language to instruct our programs to “understand” what is the connectivity information present in the spectra. For example, we know that homonuclear COSY experiments yield cross peaks that are due to 2-, 3- (and in some cases) 4-bond couplings. So a peak in a COSY peak list usually “means” either a H–C–H, H–C–C–H group is present in the molecule. The reason we are developing this is that heteronuclear experiments are still in a rapid state of development, and it is important not to “hard-code” the meaning of a cross peak into an automation program, or else you will need to rewrite the automation program when a new experiment is developed that yields cross peaks to some new coherence transfer pathway.

#### *6.1.5. Automation needs to reach results that are consistent with structural information*

One must be able to directly and interactively compare information between NMR spectra and structures. The most effective and efficient way to do this is by use of software that is capable of handling both spectra and structures. The results of the comparisons need to be available for visualization, tabulation, and for making automated decisions. Such are the capabilities of the software being developed here, and we are actively investigating the potential of this important aspect of automation.

### **6.2. Automation of structure calculations**

This is a much easier problem to solve than automation of assignments, as much of what is needed are simple routines for creating multiple structures and checking their accuracy with the original data. Also included would be stages where additional NOE or J-coupling constraints are extracted, stereospecific assignments determined, and spectra back-calculated. The most efficient means for doing this will require software that has all the necessary capabilities available. If these capabilities are in separate programs, then a means of transporting information between packages is required (the so-called seamless interface).

All of this can be done today using a variety of commercial products, where the starting structure generation, structure optimization protocols, and back calculation iterations can be automated. Future growth in this area is likely to expand the number of steps automated, and to link the structure calculations with the assignment and constraint generation tasks.

### ***6.3. Using chemical shift in assignments and structure determination***

Chemical shifts are very sensitive indicators of the structure of a molecule. The earliest structural studies of proteins using NMR focused on strongly-shifted resonances [83], which is not surprising given the technology of thirty-five years ago. However, few have tried to directly use chemical shift as a constraint in structure calculations, or to search for trends in the relationship between protein structure and chemical shifts [43,84].

Several reports have appeared where chemical shifts have been correlated with structure in proteins [47,85,86]. Very encouraging results have been obtained, and the prospects for improved fitting of chemical shift parameters to structure will be achieved given that the following occur:

1. More accurate NMR structures are obtained. Fits that use crystal structures and NMR assignments obtained under obviously different sample conditions will reduce the degree of fitting. In addition, small structure changes can have a large impact on the calculated shift.
2. Better estimations of the “random coil” values are required. For example, the current commonly-used values do not include amino acid sequence dependencies.
3. Refinements in the parameterization are made. For example, the intensity factors used in ring current shift calculations, or the description of the peptide group anisotropy contributions.

It appears that continued investigation in this direction will be fruitful.

### ***6.4. Using databases of spectra, assignment and structure data***

Considerable knowledge is present in the calculated structures and assignment tables. A database of such information that is available to all researchers is an extremely valuable asset to the NMR community.

The only example of such a database is BioMagResBank, established at the University of Wisconsin under the direction of Dr. John Markley [88,89]. The use of this database is at least twofold: it extends the amount of information available to users that otherwise could not be published in a journal article; and it allows for trends to be analyzed over a large number of spectra and structures.

It is desirable that this, or other equally accessible database be used by all in order to gain a better understanding of protein structure as well as its effects on NMR spectra. For example, the recent advances in understanding the relationship between chemical shift and protein structure is due in part through the increased availability of refined structures and corresponding resonance assignments.

### **6.5. Structure refinement and error analysis**

I expect that we will see an increase in the use of “direct” methods, where back-calculation of NOE (and J-coupling) values are incorporated into the function that is used to best fit the structure to the constraints. The benefits of this approach is the coupling of two interrelated methods (structure optimization to constraints, and constraint optimization to spectra), as well as an improvement in the final structures.

It is clear that a better means of analyzing errors in NMR structures is needed. But first we must ask: how are errors manifested in NMR structures? I believe this is the question that needs to be asked, and answered before any reasonable measure can be made of a structure's accuracy. Proteins do not exist as a single conformation in solution, so it becomes problematic to assess the accuracy of a *single* conformation, when the data clearly are obtained from a Boltzmann distribution of conformations. In addition, it is possible that the data are consistent with two or more major classes of conformations, and no method that seeks to find a single “best” structure to fit the experimental data will be satisfactory.

Furthermore, we must ask *which* constraints are to be counted in determining the agreement between structure and spectra. Are we only to consider NOEs and J-couplings when calculating the goodness of fit, or should chemical shifts be considered? Should a “good” quality structure also have good motion parameters calculated that agree with NMR relaxation data?

Perhaps a future refinement scheme will involve optimizing a structure directly to the theoretical NOE and J-couplings (as described

previously) until the errors are below a previously determined tolerance. At that point, additional potentials for optimizing against the chemical shift would be meaningful and could be introduced, as well as introduction of Lennard-Jones forces.

## 7. CONCLUSION

In the end, I suspect it will become more important to assess how well a protein structure determined by NMR helps to understand the function and physical properties of the molecule, than to know how well it represents the NMR data. If, for example, a NMR-derived structure can explain how a molecule interferes with blood clotting, is this a better measure of the "goodness" of the structure? It is fortunate that we are "cursed" with a wealth of information about molecular structure and dynamics; we need only learn how to understand the information we are presented with in the typical NMR spectrum of a protein.

## REFERENCES

1. K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.
2. S.W. Englander and A.J. Wand, *Biochemistry*, 26 (1987) 5953.
3. G.T. Montelione and G. Wagner, *J. Magn. Reson.*, 87 (1990) 183.
4. M. Ikura, L.E. Kay and A. Bax, *Biochemistry*, 29 (1990) 4659.
5. S.W. Fesik, H.L. Eaton, E.T. Olejniczak, E.R.P. Zuiderweg, L.P. McIntosh and F.W. Dahlquist, *J. Am. Chem. Soc.*, 112 (1990) 886.
6. G.M. Clore, L.E. Kay, A. Bax and A.M. Gronenborn, *Biochemistry*, 30 (1991) 12.
7. M. Ikura, L.E. Kay and A. Bax, *J. Biomolec. NMR*, 1 (1991) 299.
8. E.R.P. Zuiderweg, R. Boelens and R. Kaptein, *Biopolymers*, 24 (1985) 601.
9. S.G. Hyberts, W. Märki and G. Wagner, *Eur. J. Biochem.*, 164 (1987) 625.
10. G.T. Montelione, M.E. Winkler, P. Rauenbuehler and G. Wagner, *J. Magn. Reson.*, 82 (1989) 198.
11. P. Güntert, W. Braun, M. Billeter and K. Wüthrich, *J. Am. Chem. Soc.*, 111 (1989) 3997.
12. D.M. LeMaster and F.M. Richards, *Anal. Biochem.*, 122 (1982) 238.
13. H. Senn, B. Werner, B.A. Messerle, C. Weber, R. Traber and K. Wüthrich, *FEBS Letters*, 249 (1989) 113.
14. P.L. Weber, R. Morrison and D. Hare, *J. Mol. Biol.*, 204 (1988) 483.
15. P. Güntert, W. Braun and K. Wüthrich, *J. Mol. Biol.*, 217 (1991) 517.
16. T.F. Havel, G.M. Crippen and I.D. Kuntz, *Biopolymers*, 18 (1979) 73.

17. P.N. Barlow, M. Baron, D.G. Norman, A.J. Day, A.C. Willis, R.B. Sim and I.D. Campbell, *Biochemistry*, 30 (1991) 997.
18. Y. Boulanger, C.M. Goodman, C.P. Forte, S.W. Fesik and I.M. Armitage, *Proc. Natl. Acad. Sci. USA*, 80 (1983) 1501.
19. M. Vasák, E. Wörgötter, G. Wagner, J.H.R. Kägi and K. Wüthrich, *J. Mol. Biol.*, 196 (1987) 711.
20. M.F. Summers, T.L. South, B. Kim and D.R. Hare, *Biochem.*, 29 (1990) 329.
21. A.J. Duben and W.C. Hutton, *J. Am. Chem. Soc.*, 112 (1990) 5917.
22. P.F. Yip, *J. Magn. Reson.*, 90 (1990) 382.
23. G.M. Clore and A.M. Gronenborn, *J. Magn. Reson.*, 61 (1985) 158.
24. W. Denk, R. Baumann and G. Wagner, *J. Magn. Reson.*, 67 (1986) 386.
25. E.T. Olejniczak, R.T. Gampe Jr. and S.W. Fesik, *J. Magn. Reson.* 67 (1986) 28.
26. T.A. Holak, J.H. Prestegard and J.D. Forman, *Biochemistry*, 26 (1987) 4652.
27. I.D. Kuntz, J.F. Thomason and C.M. Oshiro, *Meth. Enzymol.*, 177 (1989) 159.
28. H. Liu, P.D. Thomas and T.L. James, *J. Magn. Reson.*, 98 (1992) 163.
29. G. Otting, H. Widmer, G. Wagner and K. Wüthrich, *J. Magn. Reson.* 66 (1986) 187.
30. D. Neuhaus, G. Wagner, M. Vasák, J.H.R. Kägi and K. Wüthrich, *Eur. J. Biochem.* 151 (1985) 257.
31. C. Redfield and C.M. Dobson, *Biochemistry*, 29 (1990) 7201.
32. L.J. Smith, M.J. Sutcliffe, C. Redfield and C. M. Dobson, *Biochem.*, 30 (1991) 986.
33. C. Griesinger, O.W. Sørensen and R.R. Ernst, *J. Am. Chem. Soc.*, 107 (1985) 6394.
34. L. Mueller, *J. Magn. Reson.*, 72 (1987) 191.
35. G.T. Montelione and G. Wagner, *J. Am. Chem. Soc.*, 111 (1989) 5474.
36. H. Kessler, C. Griesinger and K. Wagner, *J. Am. Chem. Soc.*, 109 (1987) 6927.
37. K.D. Kopple, G.R. Wiley and R. Tauke, *Biopolymers*, 12 (1973) 627.
39. M. Karplus, *J. Phys. Chem.*, 30 (1959) 11.
39. V.F. Bystrov, *Progr. Nucl. Magn. Reson. Spectros.*, 10 (1976) 41.
40. A. Pardi, M. Billeter and K. Wüthrich, *J. Mol. Biol.*, 180 (1984) 741.
41. A. Marco, M. Llinás and K. Wüthrich, *Biopolymers*, 17 (1978) 617.
42. P.E. Wright, H.J. Dyson and R.A. Lerner, *Biochemistry*, 27 (1988) 7167.
43. G. Wagner, A. Pardi and K. Wüthrich, *J. Am. Chem. Soc.*, 105 (1983) 5948.
44. J.W. Emsley, J. Feeney and L.H. Sutcliffe, *High Resolution Nuclear Magnetic Resonance Spectroscopy*. Pergamon, Oxford, 1965.
45. D.J. Ecker, T.R. Butt, J. Marsh, E.J. Sternberg, N. Margolis, B.P. Monia, S. Jonnalagadda, M.I. Khan, P.L. Weber, L. Mueller and S.T. Crooke, *J. Biol. Chem.*, 262 (1987) 14213.

46. M.A. Weiss and J.C. Hoch, *J. Magn. Reson.*, 72 (1987) 324.
47. M.P. Williamson, T. Asakura, E. Nakamura and M. Demura, *J. Biomolec. NMR*, 2 (1992) 83.
48. G.M. Clore, A.T. Brünger, M. Karplus and A.M. Gronenborn, *J. Mol. Biol.*, 191 (1986) 523.
49. R. Kaptein, E.R.P. Zuiderweg, R.M. Scheek, R. Boelens and W.F. van Gunsteren, *J. Mol. Biol.*, 182 (1985) 179.
50. K. Menger, *Jahres. Deutschen Mathem. Verein*, 40 (1931) 201.
51. T.F. Havel and K. Wüthrich, *Bull. Math. Biol.*, 46 (1984) 673.
52. T.F. Havel and K. Wüthrich, *J. Mol. Biol.*, 182 (1985) 281.
53. W. Braun and N. Go, *J. Mol. Biol.*, 186 (1985) 611.
54. P. Güntert, W. Braun and K. Wüthrich, *J. Mol. Biol.*, 217 (1991) 517.
55. P. Güntert and K. Wüthrich, *J. Biomol. NMR*, 1 (1991) 447.
56. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, *J. Comp. Chem.*, 4 (1983) 187.
57. S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner, *J. Am. Chem. Soc.*, 106 (1984) 765.
58. W.F. van Gunsteren and H.J.C. Berendsen, *Biochem. Soc. Trans.*, 10 (1982) 301.
59. P.L. Weber, D.J. Ecker, J. Marsh, S.T. Crooke and L. Mueller, *Trans. Am. Crystallogr. Assoc.*, 24 (1988) 91.
60. A.E. Torda, R.M. Scheek and W.F. van Gunsteren, *Chem. Phys. Lett.* 157 (1989) 289.
61. A.E. Torda, R.M. Scheek and W.F. van Gunsteren, *J. Mol. Biol.* 214 (1990) 223.
62. D.A. Pearlman and P.A. Kollman, *J. Mol. Biol.* 220 (1991) 457.
63. P. Yip and D.A. Case, *J. Magn. Reson.*, 83 (1989) 643.
64. A.M.J.J. Bonvin, R. Boelens and R. Kaptein, *J. Biomolec. NMR*, 1 (1991) 305.
65. J.E. Mertz, P. Güntert, K. Wüthrich and W. Braun, *J. Biomolec. NMR*, 1 (1991) 257.
66. M. Nilges, G.M. Clore and A.M. Gronenborn, *FEBS Lett.*, 219 (1987) 11.
67. I. Solomon, *Phys. Rev.* 99 (1955) 559.
68. M.J. Forster, *J. Comp. Chem.*, 12 (1991) 292.
69. M.P. Williamson, *Magn. Reson. Chem.*, 25 (1987) 356.
70. R. Boelens, T.M.G. Koning and R. Kaptein, *J. Mol. Struct.*, 173 (1988) 299.
71. B.A. Borgias and T.L. James, *J. Magn. Reson.*, 87 (1990) 475.
72. J. Fejzo, A.M. Krezel, W.M. Westler, S. Macura, J.L. Markley, *Biochem.*, 30 (1991) 3807.
73. C. Gonzalez, J.A.C. Rullman, A.M.J.J. Bonvin, R. Boelens and R. Kaptein, *J. Magn. Reson.*, 91 (1991) 659.
74. J.M. Withka, J. Srinivasan and P.H. Bolton, *J. Magn. Reson.*, 98 (1992) 611.

75. M. Billeter, V.J. Basus and I.D. Kuntz, *J. Magn. Reson.*, 76 (1988) 400.
76. C. Cieslar, G.M. Clore and A.M. Gronenborn, *J. Magn. Reson.*, 80 (1988) 119.
77. P.L. Weber and L. Mueller, *J. Magn. Reson.*, 81 (1989) 430.
78. P.L. Weber, J.A. Malikayil and L. Mueller, *J. Magn., Reson.*, 82 (1989) 419.
79. J. Xu, S.K. Straus, B.C. Sanctuary and L. Trimble, *J. Magn. Reson. Series B* 103 (1994) 53.
80. C.D. Eads and I.D. Kuntz, *J. Magn. Reson.*, 82 (1989) 467.
81. P.J. Kraulis, *J. Magn. Reson.*, 84 (1989) 627.
82. C. Eccles, M. Billeter, P. Güntert and K. Wüthrich, Abstracts 10th Meeting of the International Society of Magnetic Resonance, Morzine, France (1989) S50.
83. R. Dunkel, C.L. Mayne, J. Curtis, R.J. Pugmire and D.M. Grant, *J. Magn. Reson.*, 90 (1990) 290.
84. M. Saunders, A. Wishnia and J.G. Kirkwood, *J. Am. Chem. Soc.*, 79 (1957) 3289.
85. A Pardi, G. Wagner and K. Würthrich, *Eur. J. Biochem.* 137 (1983) 445.
86. Y. Gao, N.C. Veitch and R.J.P. Williams, *J. Biomolec. NMR*, 1 (1991) 457.
87. K. Ösapay and D.A. Case, *J. Am. Chem. Soc.*, 113 (1991) 9436.
88. P.J. Kraulis, G.M. Clore, M. Nilges, T.A. Jones, G. Petterson, J. Knowles and A.M. Gronenborn, *Biochem.*, 28 (1989) 7241.
89. E.L. Ulrich, J.L. Markley and Y. Kyoguku, *Prot. Seq. Data Anal.*, 2 (1989) 23.
90. B.R. Seavey, E.A. Farr, W.M. Westlar and J.L. Markley, *J. Biomolec. NMR*, 1 (1991) 217.

## Chapter 5

# Protein Dynamics as Studied by Solution NMR Techniques

LINDA K. NICHOLSON, LEWIS E. KAY and DENNIS A. TORCHIA

### ABSTRACT

Recent advances in the field of high resolution nuclear magnetic resonance (NMR) spectroscopy have provided a means by which the structure and dynamics of macromolecules in solution can be characterized at atomic resolution. The focus of this review is on recently developed two dimensional (2D) heteronuclear NMR techniques for measurement of nuclear spin relaxation of proteins in solution, and on the interpretation of these measurements to yield a detailed description of global and internal motions. The application of some of these techniques to staphylococcal nuclease (SNase) in the presence and absence of ligands is included to illustrate the advantages and limitations of such studies, and to demonstrate the types of information that may be obtained.

### 1. INTRODUCTION

In both macroscopic and microscopic worlds, indeed at all levels of measure, motions play a key role in the performance of functions and in the utilization of energy to perform such functions. The Victorian inventor Nikola Tesla, when remarking on the ubiquitous nature of energy and how this energy might be tapped to drive man-made machines, said in a lecture delivered before the Institution of Electrical Engineers in 1892: "Throughout space there is energy. Is this energy static or kinetic? If static, our hopes are in vain; if kinetic — and this



we know it is, for certain — then it is a mere question of time when men will succeed in attaching their machinery to the very wheelwork of nature...” [1]. Tesla’s goal at the time was to harness the power he perceived as being obtainable at any point in the universe to perform useful work in the macroscopic world. To accomplish this, he based his designs on the inherent rotation of electromagnetic energy fields, giving birth to a revolutionary class of machines powered by alternating current. Tesla possessed a rare gift of vision, not only in the field of electromagnetism, but in recognizing common themes throughout all aspects of the natural world. Were he present today, he would doubtlessly be intrigued by the field of biophysics, and by the molecular world being brought into view by modern elucidation techniques.

At the microscopic level, it is well established that atoms undergo a great deal of motion [2], and that the motion of charged particles such as electrons and nuclei generate electromagnetic fields. Biological macromolecules such as proteins appear to be naturally attached to the “wheelwork of nature” envisioned by Tesla, and their incessant motions are indeed driven by a ubiquitous energy source. If we hope to fully understand how molecular processes work, and how intramolecular communication is achieved, it is no longer sufficient to consider only structure. Although high resolution protein structure unveils the atomic scaffold in which highly specific chemical events can take place, the dynamic nature of energy and of the molecular machinery driven by it must also be brought to light.

Solution state NMR spectroscopy is a powerful tool for characterizing protein dynamics. By perturbing the state of the nuclear spin system and observing its return to equilibrium via known pathways, NMR relaxation experiments provide a sensitive means by which to probe both global and internal motions in proteins. Molecular motions generate fluctuating magnetic fields which exert a torque on the nonequilibrium magnetization, causing the system to relax back to equilibrium. Hence, the relaxation of nuclear spins is intimately tied to the motions of the atoms to which they are attached. Heteronuclei (e.g.  $^{15}\text{N}$  or  $^{13}\text{C}$ ) are well suited for relaxation studies because they provide relatively isolated spin systems for which the dominant relaxation mechanisms can be identified and isolated. In contrast, protons, which provide the bulk of information necessary for structure determination, are not well suited for relaxation studies. The interpretation of proton relaxation is complicated by rapid mutual spin flips between the observed proton and the abundant pool of neighboring protons, resulting in complex relaxation

behavior. Therefore, the heteronucleus is the preferred choice for probing dynamics, although methods have recently been proposed to measure various relaxation rates of selected protons in proteins [3]. Depending upon the types of nuclei involved and the different relaxation parameters that are measured, a wide range of timescales are accessible via solution NMR.

In order to study the dynamics of specific sites within a protein by two-dimensional heteronuclear solution NMR techniques, it is necessary to know the resonance assignments of the heteronuclei as well as their attached proton(s). Identification of each heteronuclear pair appearing in the two dimensional map to a specific site within the protein is achieved through resonance assignment techniques. Complete resonance assignments can now be obtained for proteins containing up to ~200 amino acids by combining heteronuclear labeling (incorporation of  $^{15}\text{N}$  and/or  $^{13}\text{C}$  into the protein) with recently developed double and triple resonance multidimensional techniques [4–18]. Once spectral assignments are determined, a series of two dimensional contour maps can be used to monitor the decay of magnetization of specific heteronuclei as the relaxation period is varied. In this manner the relaxation rates of specific nuclei within the protein can be determined, yielding information concerning the motions experienced by numerous sites in a protein.

To illustrate the use of some of the techniques discussed in this review, applications to SNase will be presented. Complete spectral assignments of backbone  $^{15}\text{N}$ ,  $^1\text{HN}$ ,  $^{13}\text{C}^\alpha$ ,  $^1\text{H}^\alpha$ , and selected side chains have been made [19–24]. The incorporation of  $^{13}\text{C}$  labels at selected backbone and side chain sites in SNase, including  $^{13}\text{C}^\alpha$ -Ala,  $^{13}\text{C}^\beta$ -Ala, and  $^{13}\text{C}^\delta$ -Leu labels, enable the dynamics of specific regions and types of groups within the protein to be compared. Relaxation rates in the presence and absence of ligands are presented, illustrating the dynamic response of specific regions of the protein to ligand binding. Thus, SNase is an optimal system for studying dynamics, and for characterizing the role that dynamics plays in ligand binding.

## 2. THEORY

We will begin with a discussion of the connection between nuclear spin relaxation and the random molecular reorientation experienced by a protein in solution. Next, the dominant mechanisms that provide pathways for heteronuclear spin relaxation will be described. We will

then present the equations describing the explicit dependence of observable relaxation parameters on the motional characteristics of the protein. In this brief overview we also provide references to more thorough treatments of these subjects.

### **2.1. Mathematical description of random processes**

In solution, the reorientation of a protein is governed by random collisions with solvent molecules; hence the reorientation of the protein is characterized by a broad distribution of rotational frequency components, and one cannot speak of a unique reorientation frequency or period. Mathematically the random reorientation of a protein is described by the correlation function,  $C_{kk}(t)$ , or equivalently by its Fourier transform the spectral density function,  $J_{kk}(\omega)$ .

Roughly speaking, the correlation function can be thought of as describing the orientational phase memory of an ensemble of protein molecules. For illustrative purposes we define the orientation of a protein by a vector, and consider an ensemble of vectors, all pointed along the Z axis at time  $t=0$ , with the vector sum normalized to unity. For this analogy, the correlation function is equivalent to the vector sum as a function of time. As a consequence of the random reorientation of the protein molecules the vector sum decreases and attains an equilibrium average value that depends upon the restrictions imposed upon reorientation. For isotropic reorientation the vector sum approaches zero, whereas for completely restricted motions it remains unity. Thus, the value of the correlation function at  $t = \infty$  is a measure of the restriction of reorientation. In certain cases, the NMR correlation function will decay as a single exponential function and the time constant is referred to as the correlation time. The correlation time indicates the time frame on which reorientation takes place.

Herein we restrict our considerations to axially symmetric nuclear spin interactions, which as will be seen have relevant vectors along NH and CH bond axes. For such interactions the correlation functions are given by [25,26]

$$C_{kk}(t) = \langle D_{q0}^{(2)*}[\Omega_{lab}^{k'}(t+t)] D_{q0}^{(2)}[\Omega_{lab}^k(t)] \rangle \quad (1)$$

where  $D_{q0}^{(2)}[\Omega_{lab}^k(t)]$  is a Wigner rotation matrix,  $\Omega_{lab}^k(t)$  represents the Euler angles specifying the orientation of vector  $k$  at time  $t$  relative to the laboratory reference frame, and "\*" indicates the complex conjugate

[27–29]. If  $k = k'$ , the function is dependent on the time course of a single nuclear spin interaction and is referred to as an auto-correlation function. If  $k \neq k'$ , the above equation expresses the cross-correlation between two different nuclear spin interactions  $k$  and  $k'$ , and depends upon the relative orientation of these two interactions over time.

The spectral density function can be thought of as a spectrum analyzer which expresses the probability distribution of motional frequencies experienced by the vectors under consideration, and is given by

$$J_{kk}(\omega) = \int_0^{\infty} \cos(\omega\tau) * C_{kk}(\tau) d\tau \quad (2)$$

This is directly analogous to the transformation of a free-induction-decay (FID) signal into a frequency spectrum by performing a Fourier transformation. If the correlation function is exponential in time, the spectral density function will be a Lorentzian (i.e. will have the form  $A\tau/[1+(\omega\tau)^2]$ ) centered at  $\omega = 0$  with its width determined by the correlation time  $\tau$ . For an exponentially decaying correlation function, all motions must be completely random and uncorrelated, and the spectral density function can only decrease as a function of  $\omega$ . For a given correlation time  $\tau$ , the spectral density function will have significant intensity over the frequency range  $0 \leq \omega \leq \tau^{-1}$ . As will be seen, the relaxation parameters are directly dependent upon the values of the spectral density functions evaluated at specific frequencies.

## 2.2. Relaxation mechanisms

The relaxation of a nuclear spin is governed by the fluctuations of local fields that result when molecules reorient in a strong external magnetic field [30]. The local fields experienced by a nuclear spin are a consequence of various angular dependent nuclear spin interactions [29], such as the dipolar coupling between two nuclei or the chemical shift anisotropy (CSA). These interactions can be represented by second rank tensors whose principal axes are fixed in the molecule [29]. If a nuclear spin interaction is axially symmetric, reorientation can be characterized by the motion of a single principal axis, normally taken to be the  $Z$  axis. As a consequence of molecular motion, the orientation of the  $Z$  axis (a vector) changes relative to the external field, and the strength of the local field fluctuates. The fluctuation in local field provides the mechanism for spins to exchange energy with the lattice or with each other, resulting in nuclear spin relaxation. Although

numerous relaxation mechanisms have been identified [29], we restrict our considerations to relaxation caused by the nuclear dipole–dipole and CSA mechanisms, because these interactions are the dominant sources of local fields experienced by  $^{15}\text{N}$  and  $^{13}\text{C}$  nuclei in proteins.

The dipolar interaction is a through-space coupling between two nuclear spins. The local field experienced by spin A associated with this interaction in the case of two spin 1/2 nuclei is given by

$$H_{\text{loc}} = \frac{\gamma_X \hbar}{r_{AX}^3} \left( \frac{3 \cos^2 \theta_{AX} - 1}{2} \right) \quad (3)$$

where  $\gamma_X$  denotes the gyromagnetic ratio of nucleus X,  $2\pi\hbar$  is Planck's constant,  $r_{AX}$  is the internuclear distance between A and X, and  $\theta_{AX}$  is the angle between the AX internuclear vector and the applied static magnetic field.

The CSA interaction is due to the distribution of electrons surrounding the nucleus, and the local magnetic field generated by these electrons as they precess under the influence of the applied static magnetic field. This local magnetic field tends to oppose the direction of the applied field, and hence produces a shielding effect at the nucleus. The effective field at the nucleus, taking into account this shielding due to the CSA interaction, is given by

$$H_{\text{loc}} = H_0(1 - \sigma) \quad (4)$$

where  $H_0$  is the strength of the applied static magnetic field, and  $\sigma$  is the orientationally dependent component of the CSA tensor which is given by

$$\sigma(\alpha, \beta) = \sigma_{11} \cos^2 \alpha \sin^2 \beta + \sigma_{22} \sin^2 \alpha \cos^2 \beta + \sigma_{33} \cos^2 \beta \quad (5)$$

where  $\alpha$  and  $\beta$  are the Euler angles expressing the orientation of the CSA tensor relative to the laboratory frame, and  $\sigma_{ii}$  represent the principal components, or diagonal elements of the diagonalized form, of the CSA tensor. For an axially symmetric CSA tensor in which  $\sigma_{11} = \sigma_{22} = \sigma_{\perp}$  (the component perpendicular to the symmetry axis) and  $\sigma_{33} = \sigma_{\parallel}$  (the component parallel to the symmetry axis), the above expression can be simplified to

$$\sigma(\theta_{\parallel}) = \frac{\sigma_{\parallel}}{2} (3 \cos^2 \theta_{\parallel} - 1) \quad (6)$$

where  $\theta_{\parallel}$  is the angle between the unique tensor component  $\sigma_{\parallel}$  (the component along the symmetry axis) and the applied magnetic field. Note that the maximum chemical shift will occur at  $\theta_{\parallel} = 0$  ( $\sigma = \sigma_{\parallel}$ ) and the minimum at  $\sigma_{\parallel} = 90^\circ$  ( $\sigma = \sigma_{\perp}$ ). Hence, the variation in local field due to the CSA interaction as the nucleus reorients will depend not only upon the magnitude of the applied field  $H_0$ , but also on the magnitude of the anisotropy of the CSA interaction (i.e.  $\sigma_{\parallel} - \sigma_{\perp}$ ).

A comparison of Eqs. (3) and (6) illustrates that both the dipolar and chemical shift interactions have the same orientational dependence. Hence, the fluctuation of local fields associated with these nuclear spin interactions due to the motion of the nucleus are inherently correlated. It is this correlation between field fluctuations that leads to interference effects that can substantially contribute to relaxation. Cross correlation between nuclear spin interactions must either be eliminated by experimental design or be theoretically accounted for when interpreting the relaxation data.

### 2.3. Relaxation parameters

Although numerous relaxation parameters can be measured by heteronuclear pulsed NMR experiments, the longitudinal relaxation time ( $T_1$ ), transverse relaxation time ( $T_2$ ), and the nuclear Overhauser effect (NOE) are the most straightforward to analyze. For a heteronucleus with spin  $I = 1/2$  (designated A) directly bonded to N protons (designated X) relaxed by both dipolar and CSA interactions, but in the absence of any cross correlations, the theoretical expressions for  $T_1$ ,  $T_2$  and NOE are as follows [31]:

$$\frac{1}{N*T_1} = d^2 \left[ J(\omega_X - \omega_A) + 3J(\omega_A) + 6J(\omega_X + \omega_A) \right] + c^2 J(\omega_A) \quad (7)$$

$$\begin{aligned} \frac{1}{N*T_2} = & \left( \frac{1}{2} \right) d^2 \left[ 4J(0) + J(\omega_X - \omega_A) + 3J(\omega_A) + 6J(\omega_X) + 6J(\omega_X + \omega_A) \right] + \\ & \left( \frac{1}{6} \right) c^2 [4J(0) + 3J(\omega_A)] \end{aligned} \quad (8)$$

$$\text{NOE} = 1 + \left( \frac{\gamma_X}{\gamma_A} \right) d^2 \left[ 6J(\omega_X + \omega_A) - J(\omega_X - \omega_A) \right] N*T_1 \quad (9)$$

where  $\omega_i$  denotes the Larmor frequency of nuclear species  $i$  (in rad/s),  $J(\omega)$  denotes the spectral density function evaluated at angular frequency  $\omega$ ,  $\gamma_i$  denotes the gyromagnetic ratio of nuclear species  $i$ . The constants  $d$  and  $c$  express the strength of the static dipolar and CSA interactions, respectively, and are given by

$$d^2 = \left( \frac{1}{10} \right) \frac{\gamma_A^2 \gamma_X^2 \hbar^2}{(r_{AX}^3)^2} \quad (10)$$

$$c^2 = \left( \frac{2}{15} \right) [\gamma_A H_0 (\sigma_{\parallel} - \sigma_{\perp})]^2 \quad (11)$$

where  $\hbar$  is Planck's constant ( $h$ ) divided by  $2\pi$  ( $1.0546 \times 10^{-27}$  erg·s),  $r_{AX}$  is the AX bond length ( $\sim 1.02 \times 10^{-8}$  cm for NH and  $\sim 1.09 \times 10^{-8}$  cm for CH),  $\gamma_i$  is as defined above ( $\gamma_H = 2.675198 \times 10^4$  rad·s $^{-1}$ ·Gauss $^{-1}$ ,  $\gamma_C = 0.672691 \times 10^4$  rad·s $^{-1}$ ·Gauss $^{-1}$ , and  $\gamma_N = -0.27116 \times 10^4$  rad·s $^{-1}$ ·Gauss $^{-1}$ ),  $H_0$  is the strength of the static magnetic field strength (in Gauss), and  $(\sigma_{\parallel} - \sigma_{\perp})$  is the anisotropy of an axially symmetric CSA interaction (in ppm). Calculation of these constants is simplified by using cgs units (e.g. Gauss, cm, s, rad), and by keeping all frequencies in angular units (rad/s). For the case of  $A = {}^{15}\text{N}$ ,  $\Delta\sigma = (\sigma_{\parallel} - \sigma_{\perp}) = -160$  ppm, and CSA contributions to relaxation must be taken into account. However, for  $A = {}^{13}\text{C}$ , the strength of the CSA interaction is often much smaller than that of the AX dipolar interaction in the case of aliphatic carbons, and can be neglected. For example, a rotating methyl group has an averaged anisotropy of  $\Delta\sigma \leq 25$  ppm, while a backbone alpha carbon typically has an anisotropy of  $\Delta\sigma \cong 34$  ppm. At a field strength of 11.7 Tesla, this corresponds to a contribution to relaxation of less than 2.4% and less than 2% for a rapidly rotating methyl carbon and an alpha carbon, respectively, relative to the contribution of the dipolar interaction. Finally, we note that chemical exchange sometimes contributes to the linewidth and affects the  $T_2$  measurement [32,33]. Exchange rate constants associated with slow chemical exchange can be determined from the dependence of relaxation times in the rotating frame ( $T_{1\rho}$ ) upon spin-lock field strength [34]. This method has recently been applied to uniformly  ${}^{15}\text{N}$  enriched bovine pancreatic trypsin inhibitor (BPTI) to characterize conformational rate processes [35].

It is clear from Eqs. (7)–(9) that the relaxation parameters are sensitive functions of the internuclear distance, as a consequence of the  $r^{-6}$  dependence of the term arising from the dipolar interaction. There-

fore, the interpretation of the relaxation data will be extremely sensitive to the bond length used, and the accuracy of the resulting parameters will be limited by the accuracy of the bond length.

Neutron diffraction studies of crystalline amino acids and peptides provide the best measure of bond lengths. For example, Table 1 lists the  $C^\alpha-H^\alpha$  and methyl group C-H bond lengths determined by neutron diffraction for crystalline L-Ala and L-Val [36,37]. The variability of these internuclear distances with local environment and with correction for assumed rigid body motion is evident. It is conceivable that these bond lengths for proteins in solution, where flexibility in local environments is in general greater, may range over the various values found by neutron diffraction. Taking this line of reasoning, the average  $C^\alpha-H^\alpha$  and methyl C-H bond lengths across the two structures in Table 1 can be used as an indication of the range of bond lengths possible for proteins in solution. When corrections for assumed rigid body motion are applied, the average  $C^\alpha-H^\alpha$  bond length is  $1.098 \pm 0.007$  Å, and the average methyl C-H bond length is  $1.090 \pm 0.005$  Å. The extracted amplitude for internal motion can be expressed in terms of a model-independent parameter,  $S^2$ . This parameter will be linearly scaled by the strength of the dipolar interaction; hence, the range of lengths for the  $C^\alpha-H^\alpha$  and C-H bonds correspond to differences in extracted  $S^2$  values of 7.9% and 5.6%, respectively. The extremes in the methyl C-H bond

TABLE 1

Bond lengths derived from neutron diffraction studies of crystalline amino acids

L-Ala [36]			L-Val [37]		
	bond lengths (Å)			bond lengths (Å)	
	A	B		A	B
$C^\alpha-H^\alpha$	$1.093 \pm 0.002$	1.091	$C^\alpha-H^\alpha$	$1.099 \pm 0.003$	1.105
$C^\beta-H^{\beta 1}$	$1.081 \pm 0.003$	1.096	$C^{\gamma 1}-H^{\gamma 1}$	$1.070 \pm 0.007$	1.090
$C^\beta-H^{\beta 2}$	$1.082 \pm 0.003$	1.097	$C^{\gamma 2}-H^{\gamma 2}$	$1.071 \pm 0.007$	1.084
$C^\beta-H^{\beta 3}$	$1.081 \pm 0.003$	1.096	$C^{\gamma 3}-H^{\gamma 3}$	$1.076 \pm 0.005$	1.084
			$C^{\gamma 2}-H^{\gamma 1}$	$1.076 \pm 0.006$	1.091
			$C^{\gamma 2}-H^{\gamma 2}$	$1.078 \pm 0.006$	1.089
			$C^{\gamma 2}-H^{\gamma 3}$	$1.076 \pm 0.006$	1.084

A. Not corrected for rigid body motion of nonhydrogen atoms.

B. Corrected for rigid body motion of nonhydrogen atoms.



lengths listed in Table 1 (e.g. 1.084 Å and 1.097 Å) correspond to a difference in extracted  $S^2$  values of 7.4%.

### 3. EXPERIMENTAL TECHNIQUES AND DATA ANALYSIS

In this section, recently developed pulse schemes which minimize errors in the measurement of relaxation parameters are discussed, as are data analysis techniques for determining relaxation times and their associated uncertainties. We also consider two methods for extracting information about molecular dynamics from the relaxation parameters, the model-free approach of Lipari and Szabo [25,26], and the proposed approach of Peng and Wagner for mapping out the spectral density function [38].

#### 3.1. *Pulse sequences for measurement of relaxation parameters*

Two-dimensional heteronuclear pulse schemes for measuring  $T_1$ ,  $T_2$  and NOE values for AX and AX<sub>3</sub> spin systems are shown in Figs. 1 and 2, respectively. In each pulse sequence, INEPT is used to achieve proton-heteronucleus magnetization transfer, with pulses phase cycled so that only protons directly bound to a heteroatom will be observed. The detected proton magnetization is modulated by the chemical shift of the heteronucleus during the  $t_1$  evolution period, resulting in a two dimensional spectrum with proton and heteronucleus chemical shifts along the  $f_2$  and  $f_1$  axes, respectively. Sign discrimination in the  $f_1$  dimension is achieved using time proportional phase incrementation (TPPI) [39], and WALTZ decoupling of the heteronucleus is applied during the acquisition period [40]. For the case of A = <sup>15</sup>N, effective water suppression is achieved by an off-resonance DANTE sequence applied prior to the beginning of the pulse sequence (not shown) [41,42]. Alternatively, proton antiphase coherence can be spin-locked by application of a <sup>1</sup>H spin-lock pulse, which spoils magnetization of protons (e.g. in water) not bonded to the <sup>15</sup>N [43].

##### 3.1.1. *Pulse sequences for AX spin systems*

Pulse sequences that provide accurate values of  $T_1$ ,  $T_2$  and NOE values for AX spin systems in proteins [44] are shown in Fig. 1. The  $T_1$  and  $T_2$  pulse sequences (Fig. 1A and B) utilize a double-INEPT polarization-transfer scheme to obtain approximately 30-fold and 8-fold increases in



In order to suppress the effects of AX dipolar and heteronuclear CSA cross correlation, a series of  $180^\circ$  pulses is applied during the relaxation period [44,48–51]. For the  $T_1$  experiment, these pulses are applied at a rate much greater than the relaxation rate of the fastest decaying multiplet component. For the  $T_2$  experiment, the  $180^\circ$  pulses are applied at the peak of even numbered spin echoes, again at a rate much greater than the relaxation rate of the fastest decaying multiplet component. The  $180^\circ$  pulses rapidly invert the  $^1\text{H}$  spin state, and the relaxation rates of the multiplet components are averaged. Positioning of the  $^1\text{H}$   $180^\circ$  pulses at the peak of the spin echo allows the heteronuclear magnetization to spend an equal amount of time associated with each of the two multiplet components, and ensures that antiphase magnetization does not build up. In the  $T_2$  sequence, the delay  $\delta$  is chosen such that  $\delta \ll 1/(2J_{\text{AX}})$  to avoid the build-up of rapidly decaying antiphase magnetization. It should be noted that if  $\delta$  is not short enough, a serious underestimate of  $T_2$  values can result [45,52,53]. Proton decoupling during  $t_1$  is achieved by a composite  $^1\text{H}$   $180^\circ$  pulse applied at the center of the  $t_1$  evolution period. The delay  $\Delta$  in the INEPT and reverse-INEPT schemes is usually set to slightly less than  $1/(4J_{\text{AX}})$ , while the delay  $t$  is set equal to  $1/(4J_{\text{AX}})$ . This minimizes losses due to transverse relaxation with virtually no loss in transfer efficiency.

### 3.1.2. Pulse sequences for $\text{AX}_3$ spin systems

Pulse schemes for measurement of  $T_1$ ,  $T_2$  and NOE values for  $\text{AX}_3$  spin systems (Fig. 2) differ in a number of important ways from those used for AX spin systems. For  $\text{AX}_3$  spin systems, we prefer to use single-INEPT transfer rather than the double-INEPT transfer used for the AX system. The sensitivity advantage of double- over single-INEPT transfer schemes for  $\text{CH}_3$  groups is approximately  $3.3/({}^{13}\text{C}-{}^1\text{H}$  NOE) [54]. For a methyl group in a protein, the  ${}^{13}\text{C}-{}^1\text{H}$  NOE is typically ca. 2.5, resulting in only an ca. 1.3-fold sensitivity enhancement in the double-INEPT experiment [55]. In practice this small sensitivity increase may be outweighed by signal losses associated with the additional pulses and delay periods required by the double INEPT pulse scheme.

The heteronuclear pulse schemes used to measure relaxation parameters for  $\text{AX}_3$  spin systems must be designed to equalize the contributions from individual carbon transitions to observable proton magnetization. This requirement is achieved in the case of INEPT transfer by setting the delay  $\tau$  in the reverse INEPT transfer such that

$2\pi J_{\text{CH}}\tau = 0.955 \text{ rad } (54.7^\circ)$  [56,57]. Because the  $T_2$  measurement is extremely sensitive to the length of this delay [56] and because values of  $J_{\text{CH}}$  have been found to be significantly smaller for Leu methyl groups as compared with Ala methyl groups in SNase (125 and 130 Hz, respectively) [58], it is important to measure the magnitude of the  $J_{\text{CH}}$  coupling for the particular type of methyl carbon being observed, in order that  $\tau$  be set properly to ensure that accurate measurement of the relaxation rate will be achieved.

Another potential source of error in  $\text{AX}_3$  relaxation measurements arises from the fact that the proton transitions relax with different rates during the refocusing delays in magnetization transfer schemes. This again means that individual  $^{13}\text{C}$  transitions are not transferred equally

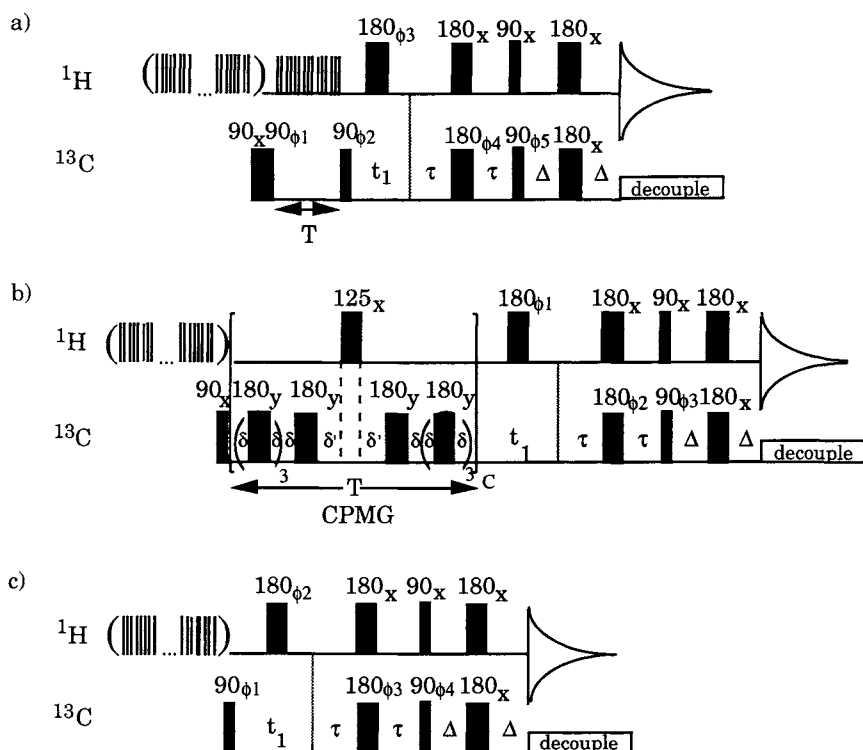


Fig. 2. Pulse schemes for the measurement of (a)  $T_1$ , (b)  $T_2$ , and (c) NOE values in  $\text{AX}_3$  spin systems. Details concerning these pulse sequences are given elsewhere [56].

to each of the individual proton transitions. In order to reduce this effect, it is desirable to minimize the magnetization transfer time. The errors introduced by this effect are smaller in the reverse-INEPT scheme than in the reverse-DEPT scheme [56]. We have found that shortening the delay  $D$  in the reverse-INEPT sequence yields relaxation parameters in slightly better agreement with those obtained by direct-observe experiments, with only small losses in sensitivity [56].

As with the AX spin system, the effects of cross correlation between dipolar and CSA interactions can be essentially eliminated by proper design of the pulse sequence. However, RF pulses cannot completely remove the effects of cross correlations between dipolar interactions. A thorough investigation of the effects of dipolar cross correlation upon methyl carbon relaxation has shown that, in general, both longitudinal and transverse relaxation processes are multiple exponential decays [59]. A recent analysis has shown that the effects of dipolar cross correlation on  $^{13}\text{C}$  longitudinal relaxation rates and on NOE values is relatively small for rapidly rotating methyl groups in proteins [60]. Using a three site hop to model internal methyl dynamics [61,62], the effects of cross correlation on longitudinal relaxation and on the NOE nearly vanish when the overall and internal correlation times are in the 5–20 ns and 15–65 ps ranges, respectively [60]. The influence of dipolar cross correlations on methyl carbon  $T_1$  measurements are also significantly attenuated when spin flips between the methyl  $^1\text{H}$ 's and neighboring  $^1\text{H}$ 's take place at a rate that is fast compared to  $1/T_1$ .

Transverse methyl carbon relaxation, however, is more strongly affected by dipolar cross correlations. The rate of mutual  $^1\text{H}$  spin flips is typically on the order of the transverse relaxation rate and cannot completely quench cross correlations. In an isolated  $\text{AX}_3$  spin system attached to a macromolecule reorienting in the  $\omega\tau_c \gg 1$  limit, the outer and inner transitions of the A quartet relax at rates that differ by a factor of 9 [60]. Hence, a rapidly rotating methyl group in a protein tumbling with  $\tau_c = 10$  ns will have two transverse decay times of approximately 23 ms and 210 ms, respectively. The application of RF pulses can only induce averaging of relaxation rates of transitions within a given manifold. For a methyl group, the relaxation rates of the four  $^{13}\text{C}$  transitions comprising the 3/2 manifold are averaged by the application of  $125^\circ$  proton pulses applied during the relaxation delay. In this case, relaxation rates associated with the 3/2 and 1/2 manifolds differ by a factor of 5, and each contribute equally to the total relaxation rate. Only proton relaxation pathways can induce mixing between the

manifolds, and the degree to which this mixing occurs will depend upon the detailed structure and dynamics of the molecule. Therefore, as a consequence of the combined effects of the  $125^\circ$  pulses and proton relaxation,  $^{13}\text{C}$  transverse magnetization decays in a biexponential manner with relaxation rates differing by less than a factor of five. The transverse relaxation time of Eq. (2) (the methyl carbon  $T_2$  in the absence of cross correlation effects) can be derived from the initial part of the transverse magnetization decay curve [59]. By reducing the differences in the relaxation rates of the various transitions, the  $125^\circ$  pulses maximize the accuracy with which the  $T_2$  values can be derived from the initial decay of the signal.

### 3.2. Data processing

Data sets can be processed using either user-written or commercially available software packages. In general, we have used NMRI software to perform the filtering, Fourier transforms and phasing of the spectra. We have used software developed by Frank Delaglio (NIH/NIDDK) [63] for determining peak volumes via a nonlinear least-squares fitting procedure, rather than the fitting routines provided in the NMRI software package. We have found that the nonlinear least-squares fitting approach yields the most consistent results, with the least sensitivity to the width used to define each peak. In this approach, cross-peak volumes are determined by simultaneously fitting the data in three dimensions, using a Gaussian function to fit each signal in the  $f_1$  and  $f_2$  dimensions, and a series of scaling factors in the time dimension. For each site, the resulting time-dependence of the peak volume is fit to a single exponential of the form  $y(t) = A_0 \exp(-t/T_i)$  using the conjugate gradient minimization technique [64]. The uncertainties of the extracted parameters,  $A_0$  and  $T_i$ , are obtained using a Monte Carlo approach described by Kamath and Shriver [65]. For this approach, the differences between the experimentally determined volumes and the best-fit exponential decay curve are used to calculate the standard deviation in volume for each site. A family of 250 synthetic data sets is then generated for each site using a Gaussian random number generator and the standard deviation calculated above to vary the volumes (e.g.  $y_j'(t_i) = y_j(t_i) + \text{sd}_j * R$ , where  $\text{sd}_j$  is the standard deviation in volume for site  $j$ ,  $R$  is a randomly generated number within a normal distribution with zero mean and unit variance, and  $t_i$  denotes experimental time point  $i$ ) and each data set is fit to an exponential function of the form

$y'(t) = A_0' \exp(-t/T_1')$ . The distribution of the parameters  $A_0'$  and  $T_1'$  thus obtained for each peak are considered an estimate of the goodness of fit for that peak.

The measurement of peak volumes in a quantitative manner must always be approached with caution, and a few practical considerations should be noted. Instrument stability, including amplifier output and receiver gain, is sensitive to environmental conditions such as temperature and humidity, as well as other factors. Sample stability may also be a factor. Reflux of the sample during long runs can occur when operating at higher temperatures, resulting in an increase of sample concentration with time. At a sample temperature of 37°C we have observed an increase in signal intensity of ~10% over a period of 3 days, with concomitant build-up of condensation on the sides of the sample tube. Field homogeneity will also be affected by the decrease in sample height caused by reflux, imparting a time dependence to linewidth which should be avoided. Measures can be taken to minimize sample reflux, such as the use of an insert in the NMR tube to prevent vapor from rising and condensing on the sides of the tube. It is advisable to repeat a number of delay times during a relaxation experiment as an assay for stability of both the instrumentation and the sample, and to interleave delay times (i.e. alternate between short and long delays during the course of the experiment) to minimize the effects of instrument and sample drift.

### ***3.3. Analysis of relaxation parameters***

The method of Lipari and Szabo [25,26] has become the commonly used approach for extracting information about protein dynamics from NMR measurements. This approach allows one to extract model-independent parameters, the generalized order parameter and effective correlation time, for internal motions in the protein, provided that the overall motion of the protein can be accurately described (e.g. isotropic vs. anisotropic tumbling). In an alternative approach, Peng and Wagner expand the number of measured NMR parameters in an attempt to directly determine values of the spectral density function in a completely model-independent fashion [38]. In this section, the advantages and disadvantages of each method will be discussed.

#### ***3.3.1. Model-free approach of Lipari and Szabo***

A protein in solution undergoes both global and internal motions, which, if on sufficiently different time scales, can be separated. In

solution, a protein tumbles globally in a random fashion, undergoing collisions with solvent molecules and with other protein molecules. If the shape of the protein is relatively spherical, global reorientation will be isotropic, having no preferred axis of reorientation. However, if the protein shape is elongated, reorientation about its long axis will be faster than about any other axis, and the global motion will be anisotropic. Anisotropic tumbling can be accounted for if necessary; for example, the relaxation behavior of sites within an ellipsoid with different rotation probabilities about its three major axes has been calculated [62,66]. For the purposes of this review, global reorientation is assumed to be isotropic.

The model-free formulation of Lipari and Szabo provides a useful way of expressing NMR spectral density functions in terms of the minimum number of motional parameters. When the overall tumbling of the protein can be described by a single correlation time and internal motions take place on a time scale much faster than the overall motion, the spectral density function is given by

$$J(\omega) = \frac{S^2\tau_m}{1 + (\omega\tau_m)^2} + \frac{(1 - S^2)\tau}{1 + (\omega\tau)^2} \quad (12)$$

where  $S$  is the generalized order parameter (a measure of the degree of spatial restriction of the motion),  $\tau_m$  is the correlation time for overall isotropic tumbling, and  $\tau^{-1} = \tau_m^{-1} + \tau_e^{-1}$ , where  $\tau_e$  is an effective correlation time characterizing the fast internal motions. It is important to note that Eq. (12) is an exact expression for the spectral density function even if multiple internal motions exist, provided that all internal motions are in the extreme narrowing motion limit (i.e.  $(\omega\tau_i)^2 \ll 1$ ) [25,26].

More complicated internal motions can also be accommodated by the model-free formalism. For the case where internal motions take place on two significantly different time scales, one of which is outside the extreme narrowing limit, the spectral density function takes the form

$$J(\omega) = \frac{S^2\tau_m}{1 + (\omega\tau_m)^2} + \frac{(1 - S_f^2)\tau_1}{1 + (\omega\tau_1)^2} + \frac{S_f^2(1 - S_s^2)\tau_2}{1 + (\omega\tau_2)^2} \quad (13)$$

where  $S_f$ ,  $S_s$ ,  $\tau_f$  and  $\tau_s$  are the generalized order parameters and effective correlation times associated with the "fast" and "slow" internal motions, respectively (i.e.  $\tau_f < \tau_s$ ),  $\tau_1^{-1} = \tau_m^{-1} + \tau_f^{-1}$ ,  $\tau_2^{-1} = \tau_m^{-1} + \tau_s^{-1}$ , and  $S^2 = S_f^2 S_s^2$  [67].

In the case of a methyl group, there are generally 2 types of motion



which contribute to relaxation. It is well established that the methyl group undergoes a 3 site hop about its three-fold symmetry axis, and we define the order parameter associated with this motion as  $S_{3R}$ . Reorientation of the symmetry axis itself will also contribute to relaxation, and we define the order parameter associated with this motion as  $S_a$ . If the correlation times associated with both of these motions are in the extreme narrowing limit, then Eq. (12) applies, with the fast internal motions represented by a single internal correlation time  $\tau_e$ , and with  $S^2 = S_a^2 S_{3R}^2$ . However, if one or both motions are characterized by correlation times that lie outside the extreme narrowing limit, the spectral density function is given by Eq. (13).

The correlation time for overall isotropic tumbling of the molecule,  $\tau_m$ , can be compared with the  $\tau_m$  value calculated for a sphere using hydrodynamic theory [68]. In the case of a sphere of volume  $V$  tumbling freely in a medium of viscosity  $\eta$ , the correlation time is given by

$$\tau_m = \frac{\eta V}{kT} \quad (14)$$

where  $k$  is Boltzmann's constant, and  $T$  is the absolute temperature. In order to obtain agreement between NMR derived values of  $\tau_m$  and those calculated using Eq. (14), it is necessary to use a 3-Å hydration shell around the protein. It should be noted that, even when the 3-Å hydration shell is included, the value of  $\tau_m$  obtained using Eq. (14) is sometimes less than the value obtained from  $^{15}\text{N}$  NMR relaxation experiments. Although the reason for this difference is not certain, it is possible that the high protein concentrations typically necessary for the NMR experiments may lead to protein aggregation or to a higher effective viscosity.

The model-free parameters associated with internal motions are of particular interest, since they are derived for each labeled site and contain detailed information about the amplitude and rate of motion of that site. As mentioned above, the effective correlation time for complex internal motion ( $\tau_e$ ) is a superposition of products of internal correlation times and geometric factors. Therefore, although  $\tau_e$  provides an estimate of the time scale of internal motion, detailed interpretation of this parameter is difficult.

In contrast with  $\tau_e$ , the generalized order parameter is directly related to the spatial extent of the internal motions even when these motions are complex. For a specific model of motion,  $S^2$  can be interpreted to yield a description of the space that would be sampled by that

motion. For example, in the case where both the A–X dipolar and the  $^{15}\text{N}$ –CSA interactions are axially symmetric with their unique tensor components coincident with the AX bond vector, the generalized order parameter is given by [25]

$$S^2 = \int d\Omega_1 p_{\text{eq}}(\Omega_1) \int d\Omega_2 p_{\text{eq}}(\Omega_2) P_2(\cos\theta_{12}) \quad (15)$$

where  $p_{\text{eq}}(\Omega_i)d\Omega_i$  is the probability at equilibrium that the unit vector  $\mu$  (along the AX bond axis) has the orientation  $d\Omega$ , where  $\Omega = (\theta, \phi)$  relative to a coordinate system fixed in the protein,  $\theta_{12}$  is the angle between  $\mu_1$  and  $\mu_2$ , and  $P_2(\cos x) = (3\cos^2 x - 1)/2$ . The order parameter can be calculated from Equation 15 for any model of reorientation of the AX bond vector. Included below (a–d) are equations for four simple but useful models of protein internal motions:

(a) Free diffusion in a cone. For free diffusion of the AX bond vector within a cone defined by the angular region  $0 < \theta < \theta_o$ ,  $0 < \phi < 2\pi$ , the order parameter is given by

$$S^2 = \left[ \left( \frac{1}{2} \right) \cos\theta_o (1 + \cos\theta_o) \right]^2 \quad (16)$$

This equation directly relates the order parameter to amplitude of motion. In general,  $S^2$  decreases as  $\theta_o$  decreases. However, it should be noted that  $S^2$  is not a monotonic function of  $\theta_o$  and has a small maximum at 120 degrees.

(b) Restricted diffusion about an axis. When motion is restricted about an axis fixed in the protein, and the AX bond vector makes an angle of  $109.5^\circ$  with this axis (corresponding to tetrahedral geometry) and diffuses through an angular range of  $\pm \gamma_o$ ,  $S^2$  is given by

$$S^2 = \left( \frac{1}{9} \right) \left[ 1 + 8 \sin^2 \gamma_o \left( \frac{1 + 2 \cos^2 \gamma_o}{3 \gamma_o^2} \right) \right] \quad (17)$$

(c) Three-fold rotation. For a methyl group executing rapid jumps between three equilibrium positions about its three-fold axis,  $S^2$  is given by

$$S^2 = \left( \frac{3 \cos^2 \beta - 1}{2} \right)^2 \quad (18)$$

where  $\beta$  is the angle between the AX bond vector and the axis of rotation.

(d) The two site jump model. If the AX bond vector abruptly changes its orientation by an angle of  $109.5^\circ$ ,  $S^2$  is given by

$$S^2 = 1 - \frac{8}{3} \left[ \frac{r}{(1+r)^2} \right] \quad (19)$$

where  $r = p_1/p_2$ , and  $p_1$  and  $p_2$  are the relative populations of the two orientations of the bond vector.

Thus, the NMR relaxation data can be interpreted using the Lipari and Szabo model-free formalism to yield generalized parameters of motion that reflect the time scale and spatial extent of internal motion. The generalized order parameter can be further interpreted for specific models of internal motion to define the region of space that would be sampled by a particular type of motion. Such information is extremely useful in visualizing the orientational fluctuations occurring within proteins, and has the potential of greatly improving the success of rational drug design approaches.

### 3.3.2. Mapping the spectral density function

An approach has been recently proposed by Peng and Wagner for directly mapping the spectral density function for an AX spin system by measuring a sufficient number of relaxation parameters [38]. Equations (7)–(9) show that the measured NMR parameters are related to the values of  $J(\omega)$  at five frequencies, i.e.  $J(\omega_H)$ ,  $J(\omega_A)$ ,  $J(\omega_H + \omega_A)$ ,  $J(\omega_H - \omega_A)$ , and  $J(0)$ . In theory, if five independent relaxation parameters could be measured, then one could directly obtain the values of  $J(\omega)$  at these five frequencies, with no assumptions concerning the nature of global or internal motions.

The three relaxation parameters discussed in Section 2 (namely,  $T_1$ ,  $T_2$  and the heteronuclear NOE) are independent of one another, and for an AX spin system are not influenced by the relaxation behavior of the directly bound proton. Peng and Wagner propose using the decay of longitudinal two-spin order ( $T_{1zz}$ ) and transverse antiphase coherence ( $T_{2AyHz}$ ) as the two additional relaxation parameters required to obtain closed form solutions for the  $J(\omega)$  terms. However, as they point out, these decay functions depend not only upon the chemical shift and dipolar interactions associated with the AX spin system, but also depend strongly upon the relaxation of the directly bound proton (the X spin) caused by dipolar interactions with neighboring protons. Peng and

Wagner approximate the spin-lattice relaxation of this proton by a single rate,  $\rho_{HH^i}$ , which is the sum of dipole–dipole relaxation rates of the H spin due to other protons  $H^i$ . Peng and Wagner report preliminary results of application of this approach to eglin c, a protein composed of 70 amino acids.

One potential problem with this approach arises from the fact that the decay of proton magnetization is multiexponential because of dipolar cross-correlations and cross-relaxation [59]. In addition, the error associated with approximating the proton relaxation as a single rate is expected to increase with molecular weight because  $\rho_{HH^i}$  increases as  $\tau_m$  increases. Hence, for larger proteins such as SNase, we prefer to use the Lipari–Szabo approach to analyze our relaxation measurements.

#### 4. EXPERIMENTAL APPLICATIONS

Staphylococcal nuclease presents many advantages for detailed dynamics studies. It is a relatively small protein (MW ~17 kD) with well characterized physical, chemical and enzymological properties [69–72]. High resolution X-ray structures [73,74] and nearly complete NMR signal assignments [19–24] are available for SNase, both in the presence and absence of ligands thymidine 3',5'-biphosphate (pdTp) and  $Ca^{2+}$ . The availability of  $^{15}N$  and  $^{13}C$  signal assignments in the liganded and unliganded states provides the opportunity to learn how ligand binding affects the internal dynamics of a protein.

The dynamics of Leu and Ala methyl side chain carbons, and Ala  $C^\alpha$  backbone sites in SNase have been investigated using pulse sequences presented in Fig. 2. Specifically labeled amino acids ([5,5'- $^{13}C_2$ ]Leucine, [2- $^{13}C$ ]Alanine and [3- $^{13}C$ ]Alanine) were biosynthetically incorporated into SNase (the two alanine labels in the same sample, the leucine in a separate sample), and relaxation parameters were measured in the absence and presence of ligands. All spectra were recorded on a Bruker AM 500 spectrometer at a temperature of 308 K on samples of 1.5 mM SNase in  $^2H_2O$  at pH 6.4 (meter reading). The solution of liganded SNase contained 5 mM pdTp and 10 mM  $Ca^{2+}$ . Data sets were recorded as  $256 \times 512$  real matrices ( $f_1 \times f_2$ ) with 16 scans per  $t_1$  point, unless otherwise noted.

A detailed analysis of the Leucine side chains in SNase has been published elsewhere [55]. Herein we will focus on the dynamics of the Ala sites, and will include the results obtained for the Leu sites for

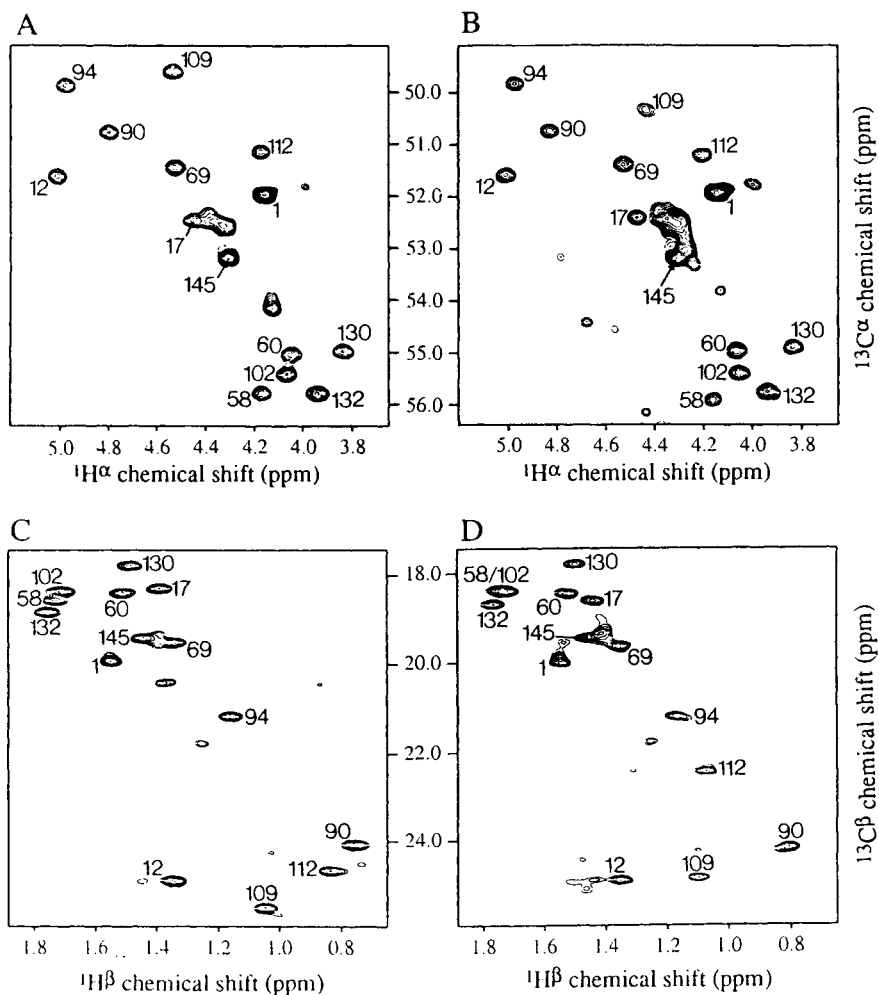


Fig. 3. Correlation spectra of  $^{13}\text{C}^\alpha$ -Ala and  $^{13}\text{C}^\beta$ -Ala labeled SNase. Each spectrum corresponds to the first plane of a set of relaxation experiments. Peaks are numbered by residue. The top two spectra represent the alpha carbon region, liganded (A) and unliganded (B). The bottom two spectra correspond to the methyl carbon region, liganded (C) and unliganded (D).

comparison. Spectra of the  $^{13}\text{C}^\alpha$ -Ala and  $^{13}\text{C}^\beta$ -Ala labeled SNase, in liganded and unliganded states are shown in Fig. 3. Residues 17, 102, 109 and 112 exhibit the largest changes in chemical shifts upon ligand binding. Except for the  $\text{C}^\beta$  signals of  $\text{A}_{58}$  and  $\text{A}_{102}$  in the spectrum of

unliganded SNase, all peaks are resolved. The large signal intensity between the  $A_1$  and  $A_{145}$   $C^\alpha$ -peaks is due to  $Ala_1$  in SNase molecules whose N-terminal Met has not been cleaved off.

The relaxation data obtained for the  $^{13}C^\alpha$ -Ala and  $^{13}C^\beta$ -Ala sites in SNase are summarized in Tables 2 and 3, respectively. Where repeated experiments were made, the average value (weighted by the number of scans) of the different measurements was used. All relaxation rates were obtained using the nonlinear least-squares approach in three dimensions to determine the decay of peak volumes as a function of time, and by fitting the decay of volumes to a single exponential using the conjugate gradient minimization approach as described in the previous section. All relaxation rate errors correspond to the precision of the mean,  $\sigma_m$ ,

$$\sigma_m = \frac{1}{\sqrt{n}} * \sigma_s \quad (24)$$

where  $\sigma_s$  is the standard deviation of the time constant as determined by the Monte Carlo approach described in the previous section, and  $n$  is the number of times the relaxation parameter was measured. The NOE values were obtained using the nonlinear least squares approach to determine peak volumes in the presence ( $V$ ) and absence ( $V'$ ) of  $^1H$  presaturation, with the NOE value given by the ratio  $V/V'$ . The error in the NOE measurement is then given by

$$\sigma_{NOE} = \frac{V}{V'} \sqrt{\left(\frac{\Delta V}{V}\right)^2 + \left(\frac{\Delta V'}{V'}\right)^2} \quad (25)$$

where  $\Delta V$  and  $\Delta V'$  denote the standard deviations in volumes  $V$  and  $V'$  respectively. To estimate  $\Delta V$  and  $V'$ , the standard deviation in baseline noise ( $\sigma$ ) was determined for each spectrum and expanded into a volume,  $\Delta V = \sigma\sqrt{j*k}$ , where  $j$  and  $k$  are the number of points in the  $f_1$  and  $f_2$  dimensions, respectively, of a typical peak at the baseline.

In order to analyze the relaxation data in terms of the minimum number of unique motional parameters, the model-free approach of Lipari and Szabo [25,26] was employed. Before the residue-specific model free parameters  $S^2$  and  $\tau_e$  can be characterized, the correlation time for overall tumbling of the protein,  $\tau_m$ , must be determined. This was accomplished using the spectral density function given by Eq. (12) together with Eqs. (7)–(9), to obtain best fits of the measured  $T_1$ ,  $T_2$  and NOE values for each  $^{13}C^\alpha$ -Ala site over a range of  $\tau_m$  values. In the

TABLE 2

Relaxation parameters measured for  $^{13}\text{C}^\alpha$ -Ala sites in SNase

Res. no.	Liganded			Unliganded		
	T <sub>1</sub> (ms)	T <sub>2</sub> (ms)	NOE	T <sub>1</sub> (ms)	T <sub>2</sub> (ms)	NOE
1	747±26	367±40	2.10±0.10	755±12	274±30	2.21±0.09
12	915±49	28±2	1.23±0.04	983±44	25±2	1.18±0.04
17	858±48	26±3	1.48±0.09	946±70	27±3	1.20±0.14
58	836±59	28±3	1.24±0.05	814±48	25±2	1.24±0.06
60	801±34	28±2	1.17±0.03	821±44	27±2	1.23±0.05
69	849±40	28±3	1.24±0.04	867±23	28±2	1.22±0.04
90	928±48	27±3	1.23±0.05	902±46	27±3	1.21±0.05
94	903±44	26±3	1.21±0.04	918±48	24±1	1.21±0.04
102	868±52	27±3	1.21±0.04	1015±65	26±2	1.22±0.05
109	854±38	27±2	1.15±0.05	873±121	25±3	1.35±0.07
112	887±75	27±3	1.21±0.04	889±58	28±1	1.31±0.06
130	888±32	26±3	1.22±0.04	948±35	26±3	1.18±0.05
132	813±41	28±2	1.30±0.04	845±30	30±2	1.29±0.05
145	371±9	126±18	1.78±0.07	372±9	124±21	1.77±0.08

TABLE 3

Relaxation parameters measured for  $^{13}\text{C}^\beta$ -Ala sites in SNase

Res. no.	Liganded			Unliganded		
	T <sub>1</sub> (ms)	T <sub>2</sub> (ms)	NOE	T <sub>1</sub> (ms)	T <sub>2</sub> (ms)	NOE
1	998±38	316±21	2.41±0.17	977±69	217±30	2.86±0.21
12	534±20	102±8	2.98±0.33	488±27	103±9	2.31±0.28
17	506±18	101±7	2.51±0.19	526±29	107±5	2.49±0.29
58	289±10	85±5	2.58±0.20			
60	584±40	106±5	2.44±0.17	516±32	107±5	1.94±0.18
69	584±26	114±10	2.68±0.29	520±26	145±8	3.17±0.37
90	464±26	87±4	2.66±0.24	425±22	98±4	2.47±0.25
94	818±46	105±6	2.60±0.22	882±47	113±11	2.45±0.23
102	421±24	76±6	2.89±0.31			
109	649±40	86±4	2.96±0.35	475±43	84±5	2.59±0.38
112	832±67	100±12	2.15±0.24	636±78	95±10	3.12±0.51
130	680±24	106±10	2.30±0.15	650±36	112±7	2.41±0.19
132	316±20	81±5	2.44±0.20	280±10	84±7	2.24±0.20
145	556±37	184±40	2.59±0.24	512±14	338±33	2.10±0.15

fitting procedure, the value of  $\tau_m$  was varied from 1 to 15 ns in increments of 0.1 ns, and the parameters  $S^2$  and  $\tau_e$  were optimized for each site. For each increment of  $\tau_m$  the net residual over all sites  $i$  ( $\chi^2$ ) was calculated according to

$$\chi^2 = \sum_i \left[ \left( \frac{T_1^e - T_1^c}{\sigma_{T_1}} \right)^2 + \left( \frac{T_2^e - T_2^c}{\sigma_{T_2}} \right)^2 + \left( \frac{\text{NOE}^e - \text{NOE}^c}{\sigma_{\text{NOE}}} \right)^2 \right]_i$$

where the superscripts  $e$  and  $c$  denote experimental and calculated values, respectively, and  $\sigma_p$  is the experimental precision of parameter  $p$  ( $p = T_1, T_2, \text{NOE}$ ). The optimal  $\tau_m$  corresponds to the minimum value of  $\chi^2$ . The Ala<sub>1</sub> and Ala<sub>145</sub> sites, which are disordered terminal residues, were omitted from the analysis. The optimal  $\tau_m$  obtained from the 12 Ala sites in liganded SNase was  $9.3 \pm 0.5$  ns, and in unliganded SNase was  $9.6 \pm 0.5$  ns. Since the value for liganded SNase was based on more data sets, this value was weighted more heavily and an overall correlation time of  $9.4 \pm 0.5$  ns is used herein. This value is somewhat larger than that previously determined from the backbone  $^{15}\text{N}$  relaxation data, obtained from a sample containing 10%  $^2\text{H}_2\text{O}$ . It has previously been shown that, when observed NOE values are near the slow motion limit value (when  $(\omega_A \tau_m)^2 \gg 1$ ), the overall correlation time can be determined by the ratio of  $T_1/T_2$  [52]. Following this approach for all backbone  $^{15}\text{N}$  sites in SNase, an overall correlation time of  $8.3 \pm 0.5$  ns was obtained [75]. As noted, use of the  $T_1/T_2$  ratio to determine  $\tau_m$  does not take into account the effects of internal motions. Therefore, the  $^{15}\text{N}$ -Ala sites were evaluated using the  $T_1$ ,  $T_2$  and NOE values and the full analysis procedure described above, which yielded  $\tau_m = 8.6 \text{ ns} \pm 0.6 \text{ ns}$ . The difference in  $\tau_m$  derived from the  $^{15}\text{N}$  and  $^{13}\text{C}$  samples is in qualitative agreement with the viscosity difference between  $^1\text{H}_2\text{O}$  and  $^2\text{H}_2\text{O}$  (at  $25^\circ\text{C}$ , the viscosity of  $^2\text{H}_2\text{O}$  is 1.23 times greater than that of  $^1\text{H}_2\text{O}$  [76]).

Using an overall correlation time of 9.4 ns, the internal dynamics of the  $^{13}\text{C}$  labeled sites in SNase were evaluated. Equations (7)–(9) were employed to determine  $S^2$  and  $\tau_e$  using the conjugate gradient minimization technique to obtain the best fit of the relaxation data, and the precision limits of these motional parameters were determined using the Monte Carlo approach as described in the previous section. To obtain the precision limits of  $S^2$  and  $\tau_e$ , the relaxation parameters and the overall correlation time were allowed to vary according to their experimental precisions. The results for the  $^{13}\text{C}^\alpha$ -Ala and  $^{13}\text{C}^\beta$ -Ala sites in the liganded and unliganded forms of SNase are summarized in



TABLE 4  
Backbone and side chain motional parameters for Ala residues in SNase

Res. no.	Liganded				Unliganded			
	$C^\alpha$		$C^\beta$		$C^\alpha$		$C^\beta$	
	$S^2$	$\tau_e$ (ps)	$S_a^2$	$\tau_e$ (ps)	$S^2$	$\tau_e$ (ps)	$S_a^2$	$\tau_e$ (ps)
1	0.42±0.20	69±6	0.16±0.03	16±2	0.08±0.15	65±3	0.27±0.05	15±1
12	0.87±0.05	16±82	0.60±0.06	27±1	0.89±0.04	11±20	0.58±0.07	30±2
17	0.85±0.04	59±30	0.60±0.06	29±1	0.87±0.05	5±49	0.56±0.06	28±4
58	0.92±0.04	36±322	0.63±0.07	55±3	0.96±0.04	86±1324		
60	0.98±0.02	35±272	0.60±0.05	24±3	0.94±0.04	45±1714	0.56±0.05	29±2
69	0.91±0.04	32±108	0.53±0.07	25±1	0.91±0.04	24±101	0.37±0.04	29±3
90	0.87±0.04	15±41	0.71±0.05	32±2	0.89±0.05	14±4314	0.59±0.05	36±2
94	0.90±0.04	14±44	0.63±0.07	16±1	0.94±0.03	19±33	0.58±0.07	15±1
102	0.92±0.04	21±370	0.81±0.08	35±2	0.89±0.04	9±14		
109	0.95±0.03	5±1205	0.76±0.07	21±2	0.92±0.05	77±148	0.74±0.07	31±3
112	0.90±0.05	16±153	0.67±0.10	16±2	0.86±0.04	32±67	0.67±0.09	22±3
130	0.91±0.04	19±78	0.60±0.08	21±1	0.88±0.04	3±8	0.56±0.05	22±1
132	0.90±0.03	52±137	0.70±0.08	50±7	0.88±0.04	40±87	0.63±0.08	58±2
145	0.51±0.03	403±46	0.28±0.10	27±2	0.51±0.03	407±50	0.08±0.03	30±2

Table 4. Note that generalized order parameter listed for the methyl sites,  $S_a^2$ , is the order parameter associated with reorientation of the  $C^\alpha-C^\beta$  bond axis. For a methyl group (having ideal tetrahedral geometry) undergoing three-fold jumps,  $S_a^2 = S^2/0.111$  [25].

#### 4.1. Backbone sites in SNase

The order parameters listed in Table 4 show that the motions of the  $^{13}C^\alpha$ -Ala backbone sites in SNase are constrained, with the exception of the N- and C-terminal residues Ala<sub>1</sub> and Ala<sub>145</sub>. The data was not fit well for these residues in either the liganded or unliganded cases, indicating that the motions of these sites are not accounted for by an effective internal correlation time that is fast compared with the overall tumbling of the protein. The terminal residues in SNase are known to be disordered, and they evidently undergo complex segmental motions that are difficult to characterize using the model free approach. Therefore, the Ala<sub>1</sub> and Ala<sub>145</sub> residues will be omitted from our analysis. The generalized order parameters for the remaining 12  $^{13}C^\alpha$ -Ala sites are  $\geq 0.84$ , with a mean value of 0.90 for both the liganded and unliganded cases. These results are in good qualitative agreement with the backbone order parameters obtained for the  $^{15}N$  backbone sites in liganded SNase [52]. A direct comparison of the  $S^2$  values obtained for the  $^{15}N$ -Ala and  $^{13}C^\alpha$ -Ala sites in liganded SNase is shown in Fig. 4. For the  $^{15}N$  and

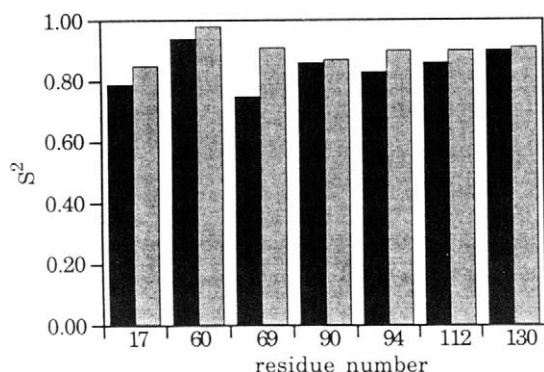


Fig. 4. Generalized order parameters determined for backbone  $^{15}N$  (black bars) and  $^{13}C^\alpha$  (light bars) Alanine sites in SNase. Overall correlation times of 8.6 ns and 9.4 ns were used in the analysis for the  $^{15}N$  and  $^{13}C$  sites, respectively.

$^{13}\text{C}$  data, overall correlation times of 8.6 ns and 9.4 ns were used, respectively. This site-specific comparison shows good agreement between the  $^{15}\text{N}$  and  $^{13}\text{C}$ -derived  $S^2$  values, indicating that the N-H and  $\text{C}^\alpha\text{-H}$  vectors associated with the same peptide plane have similar spatial restrictions to motion. Although the  $S^2$  values for the  $^{15}\text{N}$  sites are all somewhat less than their  $^{13}\text{C}$ -derived counterparts, the uncertainty in bond lengths makes this difference difficult to interpret. If one assumes that the bond vectors reorient in a cone, it is found using Eq. (16) that the observed ordered parameters correspond to cone semiangles in the range of 5–20°.

A comparison of order parameters for the  $^{13}\text{C}^\alpha\text{-Ala}$  sites in the presence and absence of ligands is shown in Fig. 5. The order parameters at each site in the liganded and unliganded forms are the same, within their precision limits, indicating that ligand binding does not change the amplitudes of the motions of the Ala  $\text{C}^\alpha\text{-H}^\alpha$  bond vectors. Since the Ala residues are dispersed throughout the protein, this result implies that the forces that define secondary structure are not significantly perturbed by the presence of the ligand. It should be noted, however, that random internal motions that occur on the same time scale as overall tumbling of the protein are difficult to separate from global reorientation. In addition, differences in significantly slower motions

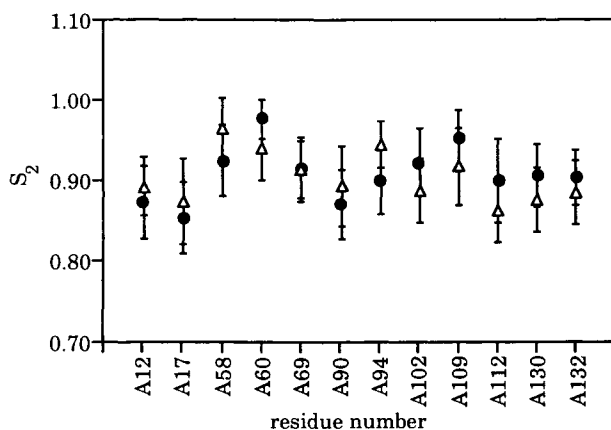


Fig. 5. Backbone  $^{13}\text{C}^\alpha\text{-Ala}$  order parameters obtained for liganded (●) and unliganded (Δ) SNase.

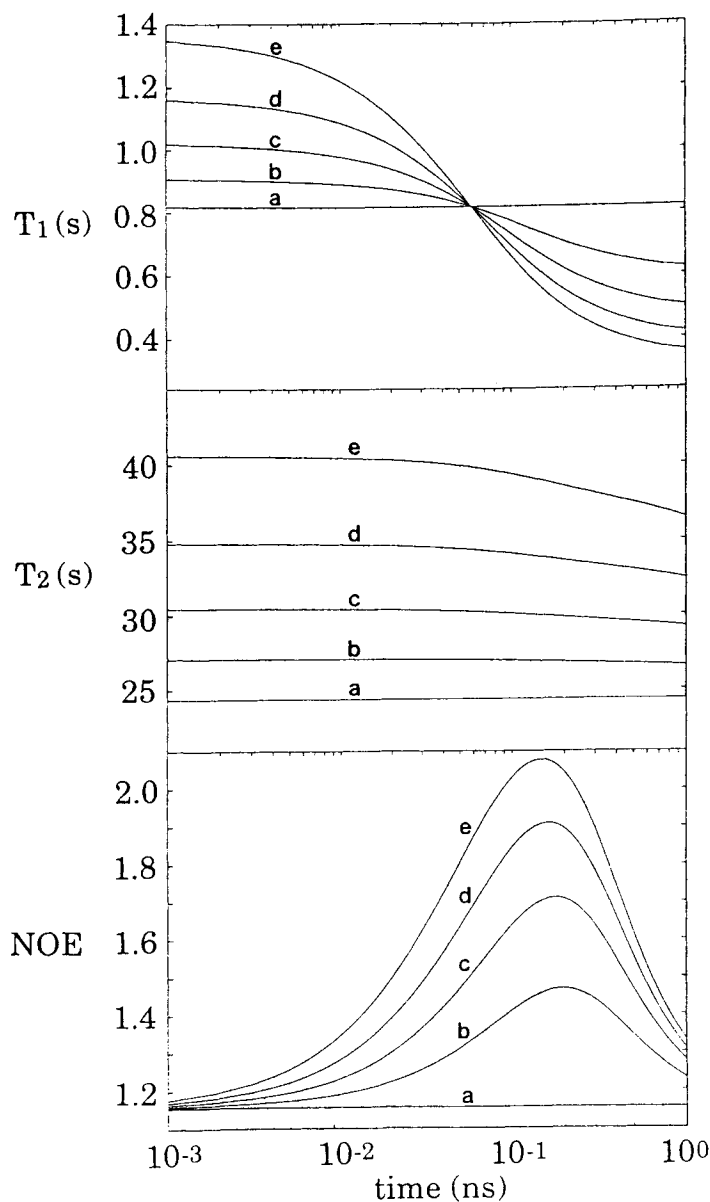


Fig. 6. Plots of Eqs. (7)–(9), which express the dependence of  $T_1$ ,  $T_2$  and NOE on  $S^2$  and  $\tau_e$ , as a function of  $\tau_e$  for various values of  $S^2$ : a = 1.0, b = 0.90, c = 0.80, d = 0.70, e = 0.60. An overall correlation time of 9.4 ns is assumed.

(e.g.  $\tau_e \gg 10$  ns) could be present, but would not be detected in the relaxation experiments.

Although the order parameters for the  $^{13}\text{C}^\alpha$ -Ala sites are well-determined, the effective correlation times vary widely from site to site, and often have very large precision limits. Inspection of Eq. (12) shows that this behavior is to be expected for highly restricted motion, since the terms that involve  $\tau_e$  becomes vanishingly small as  $S^2$  approaches unity. The relaxation parameters are plotted as a function of  $\tau_e$  for various values of  $S^2$  in Fig. 6, for an overall correlation time of 9.4 ns. It is immediately apparent that  $T_2$  is essentially independent of  $\tau_e$ , but can be used to obtain a reasonable estimate of  $S^2$ . Following curve "b" (corresponding to  $S^2=0.90$ ) in the  $T_1$  and NOE plots, it is apparent that for highly constrained motions these relaxation parameters are virtually independent of  $\tau_e$ . For example, a variation in  $\tau_e$  of nearly 2 orders of magnitude corresponds to variations of 8% and 15% respectively in  $T_1$  and in NOE. It is possible to determine accurate values of effective correlation times only when internal motions have (a) significant amplitudes and (b) correlation times that are not too small.

## 4.2. Side chain sites in SNase

The motional parameters extracted for the  $^{13}\text{C}^\beta$ -Ala sites in SNase are summarized in Table 4, where  $S_a^2$  corresponds to the spatial restriction of the  $\text{C}^\alpha$ - $\text{C}^\beta$  bond axis and  $\tau_e$  is the correlation time associated with fast internal motions, including both the 3-site hop of the methyl group and the reorientation of the  $\text{C}^\alpha$ - $\text{C}^\beta$  bond. Neglecting the unusually small values of the Ala<sub>69</sub> order parameters (which we ascribe to measurement errors resulting from the overlap of the signals of Ala<sub>69</sub> and the highly flexible Ala<sub>145</sub>) the values of  $S_a^2$  for the  $^{13}\text{C}^\beta$ -Ala sites range between 0.60 and 0.81 for liganded SNase, and between 0.56 and 0.74 for unliganded SNase. The values of  $S_a^2$ , like the values of  $S^2$  observed for the  $\text{C}^\alpha$  sites, fall into a relatively narrow range, and are relatively insensitive to the presence of ligands (Fig. 7); however, the values of  $S_a^2$  are on average only ca. 0.71 times the values of  $S^2$ . Lipari and Szabo [26] had previously noted that Ala methyl sites in BPTI had order parameters that are substantially smaller than calculated assuming that internal motion is restricted solely to three-fold rotation about the  $\text{C}^\alpha$ - $\text{C}^\beta$  bond axis. These reduced order parameters can be rationalized by assuming that the  $\text{C}^\alpha$ - $\text{C}^\beta$  bond axis librates in a cone; however, the values of the cone semi-angle required to reproduce the observed values of  $S_a^2$  are in the

range of 25–35°, i.e. substantially larger than the libration angles found for the Ala C<sup>α</sup>–H<sup>α</sup> bond axes derived from Eq. (16) and from the values of S<sup>2</sup> listed in Table 4 for the C<sup>α</sup> sites.

A possible reason that the derived values of S<sub>a</sub><sup>2</sup> are smaller than expected is that the methyl group geometry deviates slightly from tetrahedral geometry. The value of S<sub>a</sub><sup>2</sup> that one derives from S<sup>2</sup> is very sensitive to the angle β made by the C<sup>β</sup>–H<sup>β</sup> bond axis and the rotation axis (the C<sup>α</sup>–C<sup>β</sup> bond) of the methyl group. For example when β = 109.5° S<sub>a</sub><sup>2</sup> = S<sup>2</sup>/0.111 whereas when β = 111.3° S<sub>a</sub><sup>2</sup> = S<sup>2</sup>/0.91. Hence this small increase in the value of β causes derived values of S<sub>a</sub><sup>2</sup> to increase by over 20%.

It is noteworthy that single-crystal neutron diffraction studies on L-alanine [36] and L-valine [37] have shown that the values of β in these methyl groups lie in the range of 110.5–111.5°. Solid state <sup>2</sup>H NMR studies of methyl group motions in crystalline L-alanine and in alanyl peptides [77] yielded a value of β = 111.25°. It is reasonable to think that these values of β are appropriate for the Ala methyl groups in SNase, and that the values of S<sub>a</sub><sup>2</sup> listed in Table 4 should be increased by

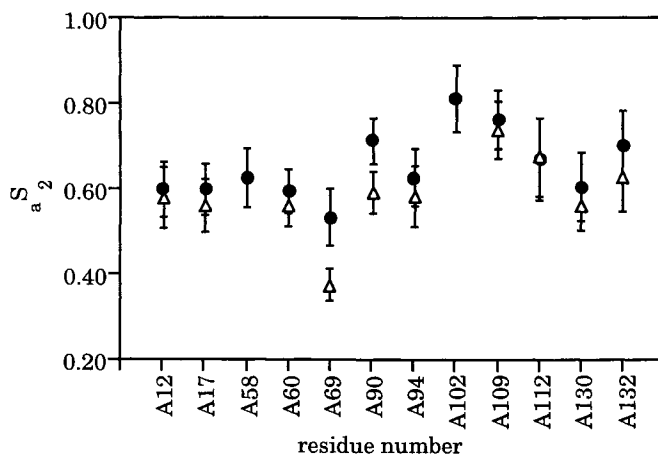


Fig. 7. Comparison of order parameters obtained for side chain C<sup>β</sup> sites of Ala residues in SNase in the presence (●) and absence (Δ) of ligands. The error bars reflect the standard deviation in each measurement as determined by a Monte Carlo analysis as described in the text.

TABLE 5

Motional parameters derived for Leucine residues in SNase

Res. no.	Liganded		Unliganded	
	$S_a^2$	$\tau_F$ (ps)	$S_a^2$	$\tau_F$ (ps)
7d1	$0.53 \pm 0.05$	$32 \pm 1$	$0.40 \pm 0.04$	$26 \pm 1$
7d2	$0.42 \pm 0.03$	$33 \pm 1$	$0.46 \pm 0.04$	$45 \pm 6$
14d1	$0.82 \pm 0.09$	$52 \pm 4$	$0.81 \pm 0.06$	$55 \pm 3$
14d2	$0.65 \pm 0.06$	$39 \pm 3$	$0.65 \pm 0.05$	$37 \pm 4$
25d1	$0.79 \pm 0.08$	$12 \pm 2$	$0.71 \pm 0.06$	$21 \pm 2$
25d2	$0.76 \pm 0.10$	$28 \pm 2$	$0.82 \pm 0.06$	$34 \pm 2$
36d1	$0.37 \pm 0.05$	$53 \pm 1$		
36d2	$0.46 \pm 0.03$	$26 \pm 1$	$0.34 \pm 0.03$	$24 \pm 1$
37d1	$0.85 \pm 0.06$	$80 \pm 3$	$0.63 \pm 0.06$	$48 \pm 2$
37d2	$0.74 \pm 0.05$	$52 \pm 1$	$0.55 \pm 0.03$	$31 \pm 2$
38d1	$0.64 \pm 0.09$	$13 \pm 4$	$0.45 \pm 0.04$	$23 \pm 1$
38d2	$0.54 \pm 0.05$	$16 \pm 4$	$0.35 \pm 0.03$	$15 \pm 1$
89d1	$0.78 \pm 0.05$	$33 \pm 2$	$0.57 \pm 0.04$	$28 \pm 2$
89d2	$0.88 \pm 0.09$	$6 \pm 1$	$0.59 \pm 0.04$	$4 \pm 1$
103d1	$0.85 \pm 0.06$	$57 \pm 2$	$0.57 \pm 0.05$	$64 \pm 2$
103d2	$0.82 \pm 0.05$	$39 \pm 3$		
108d1	$0.72 \pm 0.06$	$24 \pm 3$	$0.71 \pm 0.06$	$32 \pm 2$
108d2	$0.78 \pm 0.06$	$25 \pm 3$	$0.65 \pm 0.15$	$27 \pm 2$
125d1	$0.46 \pm 0.03$	$33 \pm 1$	$0.29 \pm 0.03$	$47 \pm 9$
125d2	$0.39 \pm 0.03$	$14 \pm 1$	$0.34 \pm 0.02$	$32 \pm 1$
137d1	$0.47 \pm 0.03$	$24 \pm 1$	$0.40 \pm 0.04$	$23 \pm 1$
137d2	$0.55 \pm 0.06$	$23 \pm 1$	$0.51 \pm 0.07$	$23 \pm 1$

15–30%, bringing them within 10–20% of the values of  $S^2$  obtained for the Ala  $C^\alpha$  sites. The remaining differences between the values of  $S_a^2$  and  $S^2$  can be accounted for by a small amplitude diffusion in a cone, having  $\theta_0$  in the range 10–20°, of the  $C^\alpha$ – $C^\beta$  bond axis.

Model free motional parameters have also been obtained for the  $^{13}C^\delta$ –Leu sites in SNase. The measurements of the Leu relaxation parameters and the procedures used to extract the model free parameters are described in detail elsewhere [55]. For the purposes of comparing the motions of the Ala and Leu methyl sites, we have repeated the analysis of the Leucine data assuming that  $\tau_m = 9.4$  ns, the same value for the overall correlation time used in the analysis of the Alanine data.

The model free parameters for the Leucine sites are listed in Table 5, where  $S_a^2$  corresponds to the spatial restriction of the  $C^\gamma-C^\delta$  bond axis and  $\tau_F$  is the correlation time associated with the 3-site hop of the methyl group. In order to facilitate a comparison of the results in Tables 4 and 5, the values of  $S_a^2$  are plotted for the Leu residues in Fig. 8. Note that the average value of the two values of  $S_a^2$  obtained for each leucine residue is plotted because the two individual values of  $S_a^2$  do not differ greatly, and plotting the average value simplifies the discussion.

Comparison of Figs. 7 and 8 show that in contrast with the values of  $S_a^2$  obtained for the Ala residues which are all close to their rms average value of 0.64, the values of  $S_a^2$  obtained for the leucine residues fall in two distinct classes. In the liganded state, leucine residues 14, 25, 37, 89, 103 and 108 have values of  $S_a^2$  that are the same,  $\langle S_{a/rms}^2 \rangle = 0.79 \pm 0.06$ , within experimental error. The value of  $\langle S_{a/rms}^2 \rangle$  for these six leucine residues is similar to that obtained for the Ala residues, and suggests that the motion of the methyl sites of these residues is as restricted as for the Ala methyl sites. The observation that the value  $\langle S_{a/rms}^2 \rangle$  of the leucine residues is about 10% larger than that of the alanine residues suggests that the angle  $\beta$  may be slightly larger in the case of the alanine residues. Leucine residues 7, 36, 125 and 137 have much

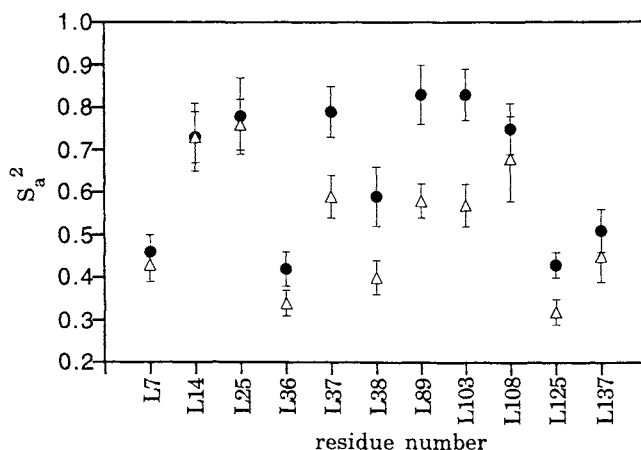


Fig. 8. Comparison of order parameters obtained for side chain  $C^\delta$  sites of Leu residues in SNase in the presence (●) and absence (△) of ligands. The error bars reflect the standard deviation in each measurement as determined by a Monte Carlo analysis as described in the text.



smaller values of  $S_a^2$ , with  $\langle S_a^2 \rangle_{\text{rms}} = 0.46 \pm 0.06$ , for these four residues. In further contrast with the behavior of the alanine methyl groups, significant differences were observed in the values of  $S_a^2$  in liganded vs nonliganded SNase for those leucine sites closest to the ligand binding site [55]. A detailed discussion of the dynamics of the leucine sites in liganded and unliganded SNase can be found elsewhere [55]. Herein we note that there is substantial variation in the amplitudes of the motions of the leucine methyl symmetry axes within both the liganded and unliganded protein molecules. In addition, the amplitudes of these motions are sensitive to the presence of ligands in the case of the leucine but not in the case of the alanine methyl groups.

Examination of Table 4 shows that the values of  $\tau_e$  are well-determined in the case of the Ala methyl groups, a result which contrasts with the poorly determined values of  $\tau_e$  found in the case of the Ala  $C^\alpha$  groups. As discussed previously, the relaxation parameters are insensitive to the value of  $\tau_e$  unless  $S^2$  is small, which is the case for methyl rotation where  $S^2$  is less than 0.1 at all sites. The values of  $\tau_e$  range from 15 to 55 ps for the Ala methyl groups, while in the case of the Leu methyl groups  $\tau_F$  (the correlation time associated with methyl rotation) ranges from 4 to 80 ps. The evident variation in  $\tau_e$  and  $\tau_F$  from residue to residue indicates that there is a significant variation in the environments of the methyl groups in the protein. Particularly striking is the seven fold difference in the  $\tau_F$  values of two methyl groups of L89. It will be of interest to see if molecular dynamics calculations can correctly predict and thereby rationalize the variations in order parameters and correlations times observed at the various backbone and sidechain sites in SNase.

## 5. CONCLUSIONS

Proton detected heteronuclear two dimensional NMR relaxation measurement techniques provide a powerful tool with which to characterize the internal motions of proteins in solution. Herein we have reviewed recently developed pulse sequences which have been optimized for measurement of  $T_1$ ,  $T_2$  and NOE values for AX and  $AX_3$  spin systems. We have illustrated the use of these techniques by application to specific backbone and side chain sites in SNase. Although in this review we have focused on a single protein, the backbone and side chain dynamics of several other  $^{15}\text{N}$  or  $^{13}\text{C}$  labeled proteins and peptides have been studied using similar techniques [67,78–86].

In general, relaxation measurements on proteins in solution require isotopic enrichment of the protein. Although  $^1\text{H}$ -detected natural abundance  $^{13}\text{C}$  relaxation studies have been successfully performed, protein concentrations of ca. 10 mM are required [87–89]. Typically, protein solubility limitations of 1–2 mM impose the requirement of isotopic enrichment levels of at least 15% to obtain the sensitivity necessary for accurate determination of relaxation parameters. Ideally, uniformly 98%  $^{13}\text{C}$  enriched proteins such as required for side chain carbon sequential assignments could be used for  $^{13}\text{C}$  relaxation studies. However, this would require the development of pulse schemes which eliminate the effects of homonuclear  $^{13}\text{C}$ – $^{13}\text{C}$  coupling on relaxation, a challenge that has not yet been met. Using current state-of-the-art pulse sequences which do not account for such couplings, uniformly 15–20%  $^{13}\text{C}$  enriched proteins should provide adequate sensitivity while reducing the effects of homonuclear coupling to an acceptable level.

A further challenge will be to develop new methods for interpreting relaxation parameters. Herein we have discussed two methods for the extraction of dynamics information, the model-free formalism of Lipari and Szabo [25,26] and the recently proposed method of Peng and Wagner for mapping out the spectral density function [38]. Although we prefer the Lipari and Szabo approach, this treatment assumes all motions are completely random, and thus does not allow for underdamped, correlated motions. Furthermore, the method yields reliable results only when internal motions are much faster than the overall motion of the protein. A more fundamental approach is to utilize molecular dynamics trajectories to obtain theoretical spectral density functions and the associated relaxation parameters, and compare these to the experimental results [90]. This comparison between theory and experiment can be used to establish the validity of the potential functions used in the molecular dynamics calculations, and to test whether or not the calculated trajectories accurately predict the internal dynamics of the protein.

Although numerous factors such as variations in bond length and local geometry contribute to the uncertainty in the parameters derived from relaxation studies, valuable information is still obtained. Site-specific comparisons between liganded and unliganded forms of a protein provide a measure of the dynamic response of the protein to the presence of the ligand. The relative motions of sites within different regions of the protein can also be compared, allowing identification of highly mobile or constrained sites. It is anticipated that future relaxation

studies will provide novel information about the internal motions of proteins, particularly at or near the protein surface, elucidating the role that dynamics plays in protein function.

## REFERENCES

1. T.C. Martin (ed.), *The Inventions, Researches and Writings of Nikola Tesla*. Barnes and Noble Books, New York, 1992.
2. M. Karplus and J.A. McCammon, *Sci. Am.*, 254 (1986) 42.
3. B. Boulat and G. Bodenhausen, *J. Biomolec. NMR*, 3 (1993) 335.
4. A. Bax, G.M. Clore and A.M. Gronenborn, *J. Magn. Reson.*, 88 (1990) 425.
5. M. Ikura, L.E. Kay and A. Bax, *Biochemistry*, 29 (1990) 4659.
6. S.W. Fesik, H.L. Eaton, E.T. Olejniczak, E.R.P. Zuiderweg, L.P. McIntosh and F.W. Dahlquist, *J. Am. Chem. Soc.*, 112 (1990) 886.
7. L.E. Kay, G.M. Clore, A. Bax and A.M. Gronenborn, *Science*, 249 (1990) 411.
8. L.E. Kay, M. Ikura and A. Bax, *J. Am. Chem. Soc.*, 112 (1990) 888.
9. L.E. Kay, *J. Magn. Reson. Series B* 101 (1993) 110.
10. G.M. Clore, L.E. Kay, A. Bax and A.M. Gronenborn, *Biochemistry*, 30 (1991) 12.
11. G.M. Clore, P.T. Wingfield and A.M. Gronenborn, *Biochemistry*, 30 (1991) 2315.
12. G.M. Clore, A. Bax, P.C. Driscoll, P.T. Wingfield and A.M. Gronenborn, *Biochemistry*, 29 (1990) 8172.
13. J.G. Pelton, D.A. Torchia, N.D. Meadow, C.Y. Wong and S. Roseman, *Biochemistry*, 30 (1991) 10043.
14. S. Grzesiek, H. Döbeli, R. Gentz, G. Garotta, A.M. Labhardt and A. Bax, *Biochemistry*, 31 (1992) 8180.
15. S. Grzesiek, J. Anglister and A. Bax, *J. Magn. Reson. Ser. B*, 101 (1993) 114.
16. S. Grzesiek and A. Bax, *J. Biomolec. NMR*, 3 (1993) 185.
17. A. Bax and S. Grzesiek, *Accounts of Chemical Research*, 26 (1993) 131.
18. S.J. Archer, A. Bax, A.B. Roberts, M.B. Sporn, Y. Ogawa, K.A. Piez, J.A. Weatherbee, M.L.-S. Tsang, R. Lucas, B.-L. Zheng, J. Wenker and D.A. Torchia, *Biochemistry*, 32 (1993) 1152.
19. D.M. Baldisseri, J.G. Pelton, S.W. Sparks and D.A. Torchia, *FEBS Lett.*, 281 (1991) 33.
20. J. Wang, D.M. LeMaster and J.M. Markley, *Biochemistry*, 29 (1990) 88.
21. J. Wang, A.P. Hinck, S.N. Loh and J.M. Markley, *Biochemistry*, 29 (1990) 102.
22. J. Wang, E.S. Mooberry, W.F. Walkenhorst and J.M. Markley, *Biochemistry* 31 (1992) 911.

23. J. Wang, A.P. Hinck, S.N. Loh, D.M. LeMaster and J.M. Markley, *Biochemistry* 31 (1992) 921.
24. D.A. Torchia, S.W. Sparks and A. Bax, *Biochemistry*, 28 (1989) 5509.
25. G. Lipari and A. Szabo, *J. Am. Chem. Soc.*, 104 (1982) 4546.
26. G. Lipari and A. Szabo, *J. Am. Chem. Soc.*, 104 (1982) 4559.
27. D.M. Brink and G.R. Satchler, *Angular Momentum*. Clarendon Press, Oxford, 1968.
28. M. Mehring, *High Resolution Spectroscopy in Solids*, 2nd Edn. Springer, Berlin, 1983.
29. H.W. Spiess, in: *NMR: Basic Principles and Progress*, Vol. 15, (P. Diehl, E. Fluck and R. Kosfeld, eds.). Springer-Verlag, New York, 1978.
30. N. Bloembergen, E.M. Purcell and R.V. Pound, *Phys. Rev.*, 73 (1948) 679.
31. A. Abragam, *The Principles of Nuclear Magnetism* (W.C. Marshall and D.H. Wilkinson, eds.). Clarendon Press, Oxford, 1961.
32. J.I. Kaplan and G. Fraenkel, *NMR of Chemically Exchanging Systems*. Academic Press, New York, 1980.
33. J. Sandström, *Dynamic NMR Spectroscopy*. Academic Press, New York, 1982.
34. C. Deverell, R.E. Morgan and J.H. Strange, *Mol. Phys.*, 18 (1970) 553.
35. T. Szyperski, P. Luginbuhl, G. Otting, P. Guntert and K. Wüthrich, *J. Biomolec. NMR*, 3 (1993) 151.
36. M.S. Lehmann, T.F. Koetzle and W.C. Hamilton, *J. Am. Chem. Soc.*, 94 (1972) 2657.
37. T.F. Koetzle, L. Golic, M.S. Lehmann, J.J. Verbist and W.C. Hamilton, *J. Chem. Phys.*, 60 (1974) 4690.
38. J.W. Peng and G. Wagner, *J. Magn. Reson.* 98 (1992) 308.
39. D. Marion and K. Wüthrich, *Biochem. Biophys. Res. Commun.*, 113 (1983) 967.
40. A.J. Shaka, J. Keeler and R. Freeman, *J. Magn. Reson.*, 53 (1983) 313.
41. A.G. Morris and R. Freeman, *J. Magn. Reson.*, 29 (1978) 433.
42. L.E. Kay, D. Marion and A. Bax, *J. Magn. Reson.*, 84 (1989) 72.
43. B.A. Messerle, G. Weder, G. Otting, C. Weber and K. Wüthrich, *J. Magn. Reson.*, 85 (1989) 608.
44. L.E. Kay, L.K. Nicholson, F. Delaglio, A. Bax and D.A. Torchia, *J. Magn. Reson.*, 97 (1992) 359.
45. A. Bax, S.W. Sparks and D.A. Torchia, *Methods Enzymol.*, 176 (1989) 134.
46. A. Bax, M. Ikura, L.E. Kay, D.A. Torchia and R. Tschudin, *J. Magn. Reson.*, 86 (1989) 304.
47. G. Bodenhausen and D.J. Ruben, *Chem. Phys. Lett.*, 69 (1980) 185.
48. M. Goldman, *J. Magn. Reson.*, 60 (1984) 437.
49. J. Boyd, U. Hommel and I.D. Campbell, *Chem. Phys. Lett.*, 175 (1990) 477.
50. A.G. Palmer, N.J. Skelton, W.J. Chazin, P.E. Wright and M. Rance, *Mol. Phys.*, 75 (1992) 699.

51. A. J. Skelton, A.G. Palmer, M. Akke, J. Kördel, M. Rance and W.J. Chazin, *J. Magn. Reson. Series B*, 102 (1993) 253.
52. L.E. Kay, D.A. Torchia and A. Bax, *Biochemistry*, 28 (1989) 8972.
53. J.W. Peng, V. Thenabal and G. Wagner, *J. Magn. Reson.*, 95 (1991) 421.
54. D.P. Burum and R.R. Ernst, *J. Magn. Reson.*, 39 (1980) 163.
55. L.K. Nicholson, L.E. Kay, D.M. Baldisseri, J. Arango, P.E. Young, A. Bax and D.A. Torchia, *Biochemistry*, 31 (1992) 5253.
56. L.E. Kay, T.E. Bull, L.K. Nicholson, C. Griesinger, H. Schwalbe, A. Bax and D.A. Torchia, *J. Magn. Reson.*, 100 (1992) 538.
57. A.G. Palmer, P.E. Wright and M. Rance, *Chem. Phys. Lett.* 185 (1991) 41.
58. G. Vuister and A. Bax, personal communication.
59. L.G. Werbelow and D.M. Grant, *Adv. Magn. Reson.*, 9 (1977) 189.
60. L.E. Kay and D.A. Torchia, *J. Magn. Reson.*, 95 (1991) 536.
61. D.E. Woessner, *J. Chem. Phys.*, 36 (1962) 1.
62. D.E. Woessner, *J. Chem. Phys.* 37 (1962) 647.
63. F. Delaglio and A. Bax, personal communication.
64. W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes in C*. Cambridge University Press, Cambridge, 1988.
65. U. Kamath and G.W. Shriver, *J. Bio. Chem.*, 264 (1989) 5586.
66. D.E. Woessner, *J. Chem. Phys.*, 37 (1962) 647.
67. G.M. Clore, A. Szabo, A. Bax, L.E. Kay, P.C. Driscoll and A.M. Gronenborn, *J. Am. Chem. Soc.*, 112 (1990) 4989.
68. R.M. Venable and R. W. Pastor, *Biopolymers*, 27 (1988) 1001.
69. P.W. Tucker, E.E. Hazen and F.A. Cotton, *Molec. and Cell. Biochem.*, 22 (1978) 67.
70. P.W. Tucker, E.E. Hazen and F.A. Cotton, *Molec. and Cell. Biochem.*, 23 (1979) 3.
71. P.W. Tucker, E.E. Hazen and F.A. Cotton, *Molec. and Cell. Biochem.*, 23 (1979) 67.
72. P.W. Tucker, E.E. Hazen and F.A. Cotton, *Molec. and Cell. Biochem.*, 23 (1979) 131.
73. P.J. Loll and E.E. Lattman, *Proteins: Structure, Function, and Genetics*, 5 (1989) 183.
74. T.R. Hynes and R.O. Fox, *Prot. Struct. Funct.*, 10 (1991) 92.
75. D.A. Torchia, L.K. Nicholson, H.B.R. Cole and L.E. Kay, in: *Topics in Molecular and Structural Biology Series: NMR of Proteins* (G.M. Clore and A.G. Gronenborn, eds.). MacMillan, in press.
76. P.G. Stecher, M. Windholz and D.S. Leahy (eds.), *The Merck Index*, Eighth Edn. Merck & Co., Rahway, NJ, 1968.
77. L.S. Batchelder, C.H. Niu and D.A. Torchia, *J. Am. Chem. Soc.* 105 (1983) 2228.
78. J.W. Peng, V. Thenabal and G. Wagner, *J. Magn. Reson.*, 94 (1991) 82.
79. G. Barbato, M. Ikura, L.E. Kay, R. Pastor and A. Bax, *Biochemistry*, 31

- (1992) 5269.
80. D.M. Schneider, M.J. Dellwo and A.J. Wand, *Biochemistry*, 31 (1992), 3645.
  81. M.J. Stone, W.J. Fairbrother, A.G. Palmer, J. Reizer, M.H. Saier and P.E. Wright, *Biochemistry*, 31 (1992) 4394.
  82. H. Takahashi, E. Suzuke, I. Shimada and Y. Arata, *Biochemistry*, 31 (1992) 2464.
  83. M. Akke, N.J. Skelton, J. Kördel, A.G. Palmer and W.J. Chazin, *Biochemistry*, 32 (1993) 9832.
  84. J. Kördel, N.J. Skelton, M. Akke, A.G. Palmer and W.J. Chazin, *Biochemistry*, 31 (1992) 4856.
  85. A.G. Palmer, R.A. Hochstrasser, D.P. Millar, M. Rance and P.E. Wright, *J. Am. Chem. Soc.*, 115 (1993) 6333.
  86. B.L. Grasberger, A.M. Gronenborn and G.M. Clore, *J. Molec. Biol.*, 230 (1993) 364.
  87. N.R. Nirmala and G. Wagner, *J. Magn. Reson.*, 82 (1989) 659.
  88. M.J. Dellwo and A.J. Wand, *J. Am. Chem. Soc.*, 111 (1989) 4571.
  89. A.G. Palmer, M. Rance and P.E. Wright, *J. Am. Chem. Soc.*, 113 (1991) 4371.
  90. A.G. Palmer and D. Case, *J. Am. Chem. Soc.*, 114 (1992) 9059.

## Chapter 6

# Nucleic Acid Structure and Dynamics from NMR

DAVID E. WEMMER

### 1. INTRODUCTION

The power of multidimensional NMR methods has opened new opportunities for analysis of the structure of nucleic acids as well as proteins in solution. There are detailed discussions giving both principles and applications of two and higher dimensional NMR in previous chapters. In the present chapter we focus on recent applications of this NMR methodology to nucleic acid systems. There have been a number of reviews covering somewhat older work on nucleic acids [1], hence here the focus will be on recent progress and examples. The basic methods for obtaining resonance assignments in duplex nucleic acids have been well described previously [2], and will be reviewed briefly. However a number of modifications will be described which are advantageous, particularly for more complex systems. In the following section the sources for generating direct structural information will be discussed, with a description of how they are applied in nucleic acid systems. Next the methods through which the structural information can be used to develop models for the solution structures will be presented. The NMR methods for analyzing dynamics in nucleic acid systems will also be discussed. Finally specific examples of the application of these NMR approaches will be given.

## 2. RESONANCE ASSIGNMENTS IN NUCLEIC ACIDS

The first, crucial step in any structure interpretation is assignment of resonances, that is the association of specific resonances with specific spins in the molecule. The general sequential assignment strategy for  $^1\text{H}$  resonances was developed for proteins [3], but was soon extended to other biopolymers as well. The key element in this process is the covalent connectivity of a backbone in the molecule, the sugar phosphate system in DNA and RNA. The structures of the individual residues which occur in these molecules are shown in Fig. 1. The bonds which are present connecting these residues assure that there are very frequently protons on neighboring residues which are close in space. This proximity leads to NOEs which can be detected in multidimensional spectra. The type of residue at each step can be identified through the patterns of couplings observed either in correlated spectra (COSY or TOCSY) or through NOEs, together with information from the primary sequence of the molecule. This combination of proximity and sequence can generally lead to assignment of a majority of the reso-

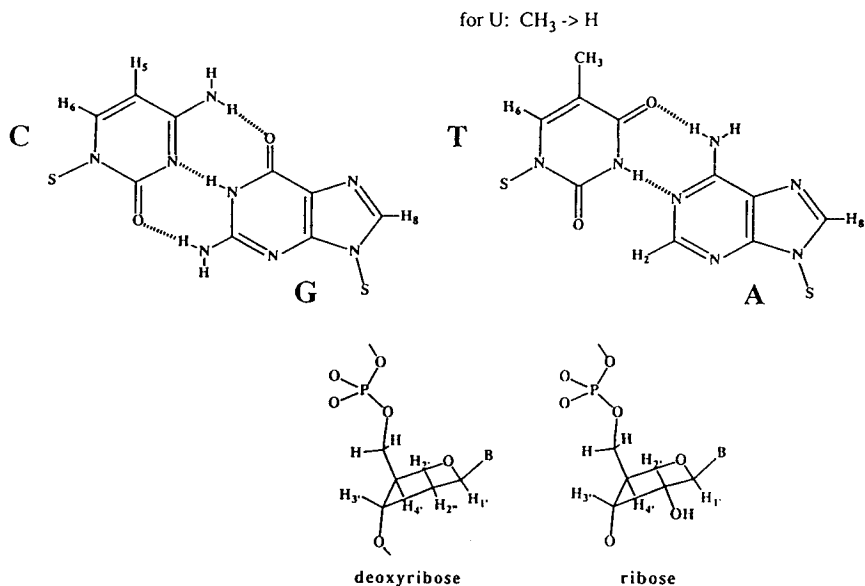


Fig. 1. Structures of the Watson-Crick base pairs occurring in nucleic acids, and the structures of the deoxyribose and ribose sugars occurring with each nucleotide.



nances. With the  $^1\text{H}$  resonances assigned, the  $^{13}\text{C}$  and  $^{31}\text{P}$  assignments follow readily using heteronuclear correlation experiments. Alternative approaches have been described in which just couplings between residues are used. For nucleic acids these necessarily involve the  $^{31}\text{P}$  of the phosphate, since there are no measurable  $^1\text{H}$ – $^1\text{H}$  interresidue couplings. However many of the relevant couplings are small, and the method does not work well in larger nucleic acids.

The chemically distinct types of protons in nucleic acids have different resonance frequencies, just as in any other organic molecule. The chemical shifts are affected by the identities of the neighboring residues, and by the conformation, but the overall character of the spectrum is still dominated by the chemical nature of the protons observed. In nucleic acids there are some protons attached to nitrogen or oxygen, which are acidic, and exchange fairly rapidly with solvent. These are the imino (in the center of a base pair), the aminos (the outer hydrogen bonds of base pairs), and the ribose hydroxyl. These can generally be seen only in  $\text{H}_2\text{O}$  solvent, but provide useful probes both for structure (especially base pairing) and dynamics. If the imino protons are not involved in a hydrogen bond, then they are usually broadened through rapid solvent exchange. The two protons of a single amino group can also exchange through rotation about the C–N bond. This is hindered through conjugation of the nitrogen's electrons into the p system of the ring. The amino protons of cytosine usually rotate fairly slowly, the two protons giving rise to separate resonances — the hydrogen bonded one generally resonating downfield of the one which is not. The aminos of adenosine and guanosine rotate at an intermediate rate, leading to severe exchange broadening. Often these cannot be seen at all, at other times they are broad lumps under other sharper peaks.

The majority of protons, however, are attached to carbon and exchange very slowly, if at all. These can be observed in  $\text{D}_2\text{O}$  solutions, and NOEs involving these are the major source of structural details. A spectrum of a DNA containing eleven base pairs is shown in Figure 2, with the spectral regions containing the different types of protons identified. RNA does not contain the deoxy sugar position, and hence there is less dispersion of the sugar resonances. This makes the assignment problem somewhat more complex, as will be described below. It is worth noting that the structure of each sugar of a DNA or RNA is chemically identical, and hence gives rise to basically the same coupling pattern. In addition to the sugars there are aromatic protons on the bases. The T and C bases in DNA are chemically distinct, and give rise

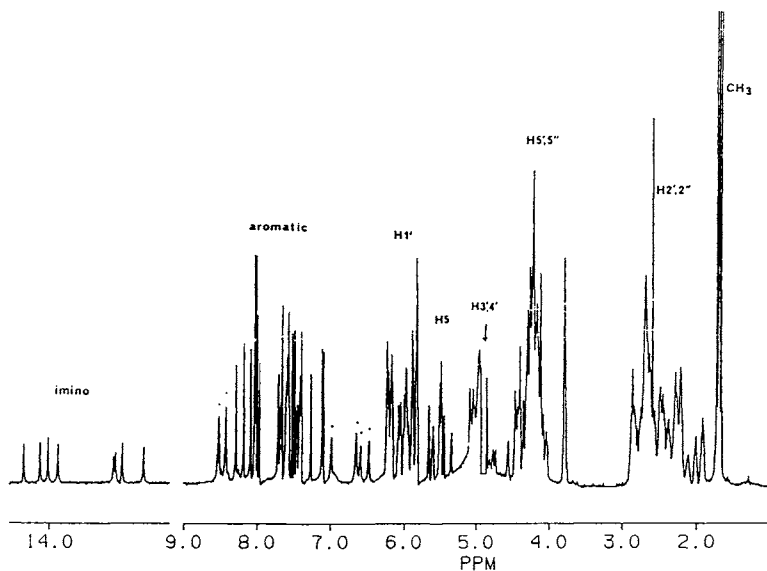


Fig. 2.  $^1\text{H}$  spectrum of the DNA duplex 5' CGCAAAGGC 3' + 5' GCCTTTGCG 3', taken in  $\text{H}_2\text{O}$  solution. The regions in which different types of protons have resonances are labelled with the proton type. The solvent resonance was suppressed with presaturation at the water frequency, indicated by the arrow.

to characteristic coupling patterns. The H5–H6 coupling in cytosine is fairly large, making each proton a doublet. The thymine H6– $\text{CH}_3$  coupling is rather small (ca. 1 Hz), but can be seen in COSY spectra for small oligomers. In any case the H6– $\text{CH}_3$  NOE is distinctive, and is usually easily identified. The H8 resonances of both guanosine and adenosine are singlets, with similar chemical shift, and hence cannot easily be distinguished. Sequence information generally must be applied to make the final identification of these, as discussed further below. Adenosine has an additional aromatic proton at the 2 position, giving another singlet. This proton is close to the imino in A:T base pairs, and can be identified through the NOE to it. In RNA thymine is replaced by uracil, in which the methyl is replaced by a proton. Thus in this case U and C give the same coupling pattern, and hence again sequence information must be used in making a final distinction between these bases.

In many cases of biological interest the segments of DNA or RNA take on a regular, right handed helical conformation. Although the

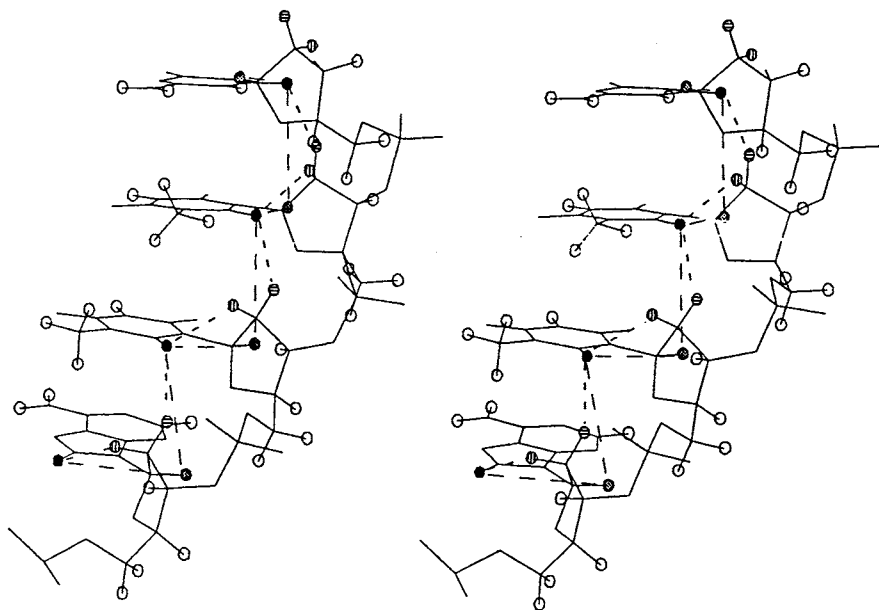


Fig. 3. A segment of DNA with the sequence GTTC is shown as a stereo pair, in the right handed helical conformation of B-DNA. The dashed lines indicate connectivities observable in NOE spectra between aromatic and sugar protons, which can be used to obtain the sequential assignments.

details in DNA are different from RNA, they have many features in common. An expanded segment of the backbone for a DNA in the idealized "B" conformation is also shown in Fig. 3. The regular repeating positions of the bases and sugars in such a structure leads to a repeating pattern of NOEs. There are two types of proton connections which are indicated,  $H_a-H_{1'}$ , and  $H_a-H_{2',2''}$  (where  $H_a$  is the aromatic H6 or H8 proton). Although some of the distances between these pairs of protons are near the limit for observable NOEs (4–5 Å), they can usually be seen in spectra. Thus using a single region of a NOESY spectrum it is possible to identify each pair of neighboring bases.

In a general DNA sequence (without symmetry) each strand gives rise to a set of connectivities. The full NOESY spectrum of a DNA oligomer is shown in Fig. 4. The  $H_a-H_{1'}$  region is shown expanded in Fig. 5, for the DNA undecamer, with each expected connectivity seen. Since these can be followed end to end there is no ambiguity in assignments. Each cytosine is easily identified in this region through the strong

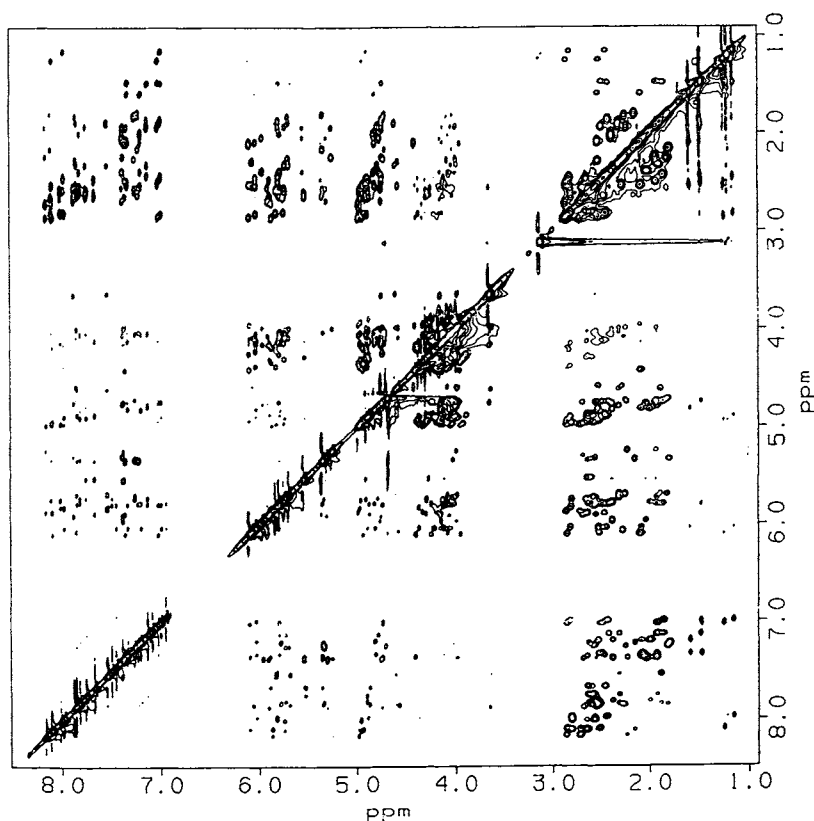


Fig. 4. A NOESY spectrum, taken in  $D_2O$  solution, of the DNA undecamer 5'CG-CAAATTGGC3' + 5'GCCAATTITGCG3' is shown. The spectral regions can be identified through the regions identified in Fig. 2.

H5–H6 cross peak, and, if the spectrum is run in  $H_2O$ , through the NOEs to the amino protons of the cytosine. The thymines are similarly identified through NOEs to the methyl protons. Since the sequence was synthesized, and the two strands are known to be complementary, the identity of the G and A residues is derived from their position in the sequence. A similar tracing can be done in the  $H_a$ – $H_{2',2''}$  region, providing an independent check on the connectivities. The process is similar for RNA, however the  $H_a$ – $H_{1'}$  NOEs are weaker (longer distances between protons), and the  $H_{2'}$  resonances are overlapped with many other sugar protons. Working in  $H_2O$  solutions it is possible to observe other

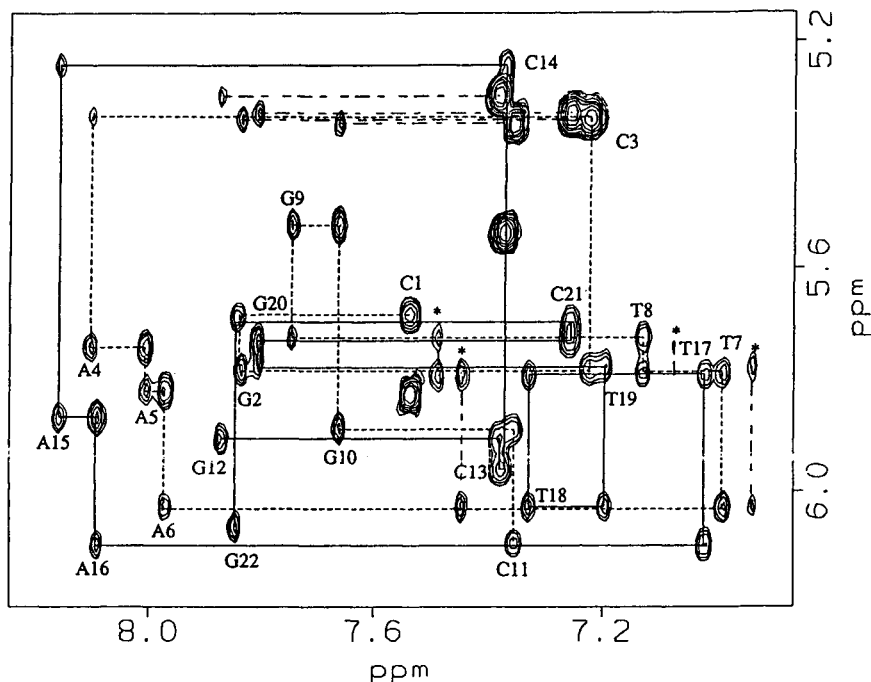


Fig. 5. An expansion from NOESY spectrum, taken in  $D_2O$  solution, of the DNA undecamer  $5'CGCAAATTGGC3' + 5'GCCAATTTGCG3'$  is shown. The horizontal axis has resonances from aromatic C6 or C8 protons, and the vertical those of sugar C1' and cytosine C5 protons. The solid line traces the connectivities from one strand, and the dashed line those from the other. Asterisks indicate the positions of adenosine C2 protons, and the dot-dashed lines the connections between aromatics and cytosine C5 protons.

connectivities involving amino protons, which also help in the assignment process [4]. For small RNAs the  $H\alpha$ - $H1'$  connectivities are sufficient to obtain assignments, but for longer sequences other tricks must be used, as discussed further below.

In longer DNA sequences there may be accidental degeneracies in chemical shifts of either base or sugar protons (or, even worse, both). In such cases the sequence information again comes into play. At such a degeneracy there are two possible connections, which are equivalent in terms of spectral consistency. However the correct set must lead to the correct overall sequence of the DNA (or RNA), hence both possibilities can be tested. Usually only one of the assignment possibilities is com-

patible with the sequence, and the degeneracy is resolved. It is important to remember that the connectivities observed in *all* possible spectral regions must be consistent with the final assignment and sequence. Although they are not visible for all base steps,  $H_a-H_a$  or  $H_a-CH_3$  connections provide important checks on the assignments. Couplings identified in COSY or TOCSY spectra similarly provide a check on which sugar protons arise from the same residue. Double quantum spectra can also be valuable, providing the same information, but in a sufficiently different format that it complements the direct methods. It is particularly valuable when there are degeneracies in resonances of coupled protons, for example the  $H_2, H_2'$ . This methodology has been applied recently to a four strand immobile DNA junction containing 64 nucleotides [5]. Although there were many degeneracies, and the resonances are quite broad, systematic application of the sequential assignment rules allowed assignment of most base and sugar resonances.

In large nucleic acids even the identification of residue types can be difficult, especially for RNA in which U and C cannot directly be distinguished from one another. There are a number of sources of information which can be used to supplement the sequential approach. Purely spectroscopic approaches can be of benefit. A preTOCSY-NOESY step, transferring magnetization from the H5 to the H6 protons, has been used in DNA triplexes to selectively discriminate signals from cytosines [6]. Another similar trick is to use a Hartman-Hahn mixing step between  $^{31}\text{P}$  and coupled protons, then to transfer this to other protons of the residue using TOCSY mixing sequences [7]. In addition  $^1\text{H}$  3D experiments have been applied to nucleic acids [8], extending resonance assignments considerably. The most used 3D experiments are generally a combination of TOCSY and NOESY. As the molecular weight increases the TOCSY transfer becomes less efficient, lowering the overall sensitivity of the experiment. In addition, since a coherent transfer step is *required* it is less favorable in RNA systems for which some of the sugar couplings are very small, particularly the  $H1'-H2'$  in  $C3'$  endo sugars typical of the A-form RNA conformation. 3D NOESY-NOESY experiments have also been done on DNA [9]. In this case there is no coherence transfer step, and hence the performance does not degrade with increasing molecular weight. However the interpretation is somewhat more complex since there are no steps which are necessarily restricted to be intraresidue.

Isotope labelling can be used to distinguish different residue types. The easiest labelling reaction to carry out is deuteration, which can be

achieved after synthesis of the nucleic acid. The purine (A and G) H8 aromatic protons will exchange with solvent  $^2\text{H}_2\text{O}$  under the proper conditions (increasing in rate at elevated pH and temperature). Cytosine and uracil will not exchange under these conditions, but if bisulfite is added an addition/elimination cycle occurs which leads to exchange of the H5 [10]. Clearly since this position has a methyl group in thymine, the same reaction cannot occur. For many nucleic acids of interest two strands are synthesized to form a duplex, and these can then be differentially labelled before combining them. Having a defined subset of resonances removed from 2D spectra can verify that these do arise from a specific base type and strand. Deuteration is basically negative labelling (removal of resonances) for  $^1\text{H}$  NMR, direct observation is not attractive since  $^2\text{H}$  has a quadrupole moment which leads to relatively broad lines in solution. However it is useful for direct observation in solids for dynamic studies in which the quadrupole coupling is an asset [11].

Another approach is to use the large chemical shift dispersion of  $^{13}\text{C}$  to separate  $^1\text{H}$  resonances by carbon type. The chemically distinct carbons, in both the bases and sugar, have chemical shift differences which are larger than those due to sequence or secondary structure. Thus in a  $^{13}\text{C}$ - $^1\text{H}$  correlation experiment the resonances are grouped according to carbon type [12]. With this knowledge about the chemical identity for the different protons, the assignment using  $^1\text{H}$  NOESY spectra is easier. These heteronuclear correlation experiments can be collected with  $^1\text{H}$  detection, giving high enough sensitivity to perform at natural abundance for nucleic acids at a few mM concentration [13].

Isotope labelling can also be done to incorporate  $^{13}\text{C}$  or  $^{15}\text{N}$  at levels above their natural abundance (1.1% and 0.3% respectively). This can be done through chemical synthesis of labelled bases or sugars (usually at a limited number of sites), followed by either chemical [14] or biochemical [15] incorporation into a nucleic acid. The resonances of protons attached to these labels can be detected in a variety of filtering experiments [16], in addition to greatly enhanced sensitivity in correlation experiments. The filters can be arranged so that on either axis of a 2D experiment only protons which are (or are not, depending on the design) attached to the heteroatom are detected. This can lead to a tremendous reduction in the number of resonances being detected, and hence simplification of detection. Recently high level, uniform labelling has been achieved for RNA [17] (using a combination of *in vivo* isotope incorporation, and *in vitro* enzymatic synthesis). This makes possible

use of the 3D and 4D experiments which have been so successful in extending the molecular weight range of proteins which could be studied. This approach has the best features of filtering and correlation experiments, i.e. both simplification and new information. Because of the high sensitivity of the correlation part of these experiments it is possible to transfer the whole NOESY spectrum from a particular proton to the  $^{13}\text{C}$  or  $^{15}\text{N}$  to which it is attached. For the sugars in RNA, with extremely crowded proton spectra, this is a major improvement. This approach will probably be required for systematic study of any large RNA (30+ nucleotides).

A final isotope label, which will probably see more limited use, is tritium,  $^3\text{H}$ . Tritons can be exchanged for protons at specific sites in nucleotides, and then incorporated into larger nucleic acids.  $^3\text{H}$  has extremely good characteristics for NMR, slightly higher sensitivity than  $^1\text{H}$  with the same chemical shift and coupling characteristics. It has the disadvantage of being radioactive, and hence requires a specialized laboratory for synthesis of samples. Fortunately it is a very weak  $\beta^-$  emitter, the radiation being completely blocked by the glass of an NMR tube or the typically used teflon liner. It is useful in direct detection experiments, and may be of particular value in cases where there is significant line-broadening and inverse detection of  $^1\text{H}$  does not work well [18]. NOEs to or from  $^1\text{H}$  can be detected with basically the same characteristic as homonuclear NOEs. These can aid in assignments in the cases where the labelling alone does not provide complete assignments.

### 3. PARAMETERS FOR STRUCTURE DETERMINATION

There are basically two direct sources of information from NMR spectroscopy, angles from coupling constants and distances from NOEs. It is well known that three bond coupling constants (either homonuclear or heteronuclear) are sensitive to the dihedral angle connecting the planes of the bonds. The value can be estimated through use of a Karplus relation, an equation relating the coupling constant to angle. For example:

$$^3J(\text{HCOP}) = 15.3 \cos^2(\phi) - 6.1 \cos(\phi) + 1.6 \text{ Hz}$$

where  $\phi$  is the dihedral angle. There are a number of constants which appear in such equations which must be experimentally determined.



This is done through measurement on compounds for which the structure is known and there is little or no conformational flexibility. In nucleic acids coupling values provide useful information about the conformation (pseudorotation) of the sugar ring, and about the angles connecting the sugars to the phosphates. The calibration of the Karplus relation for the hydrogens in the sugar rings has been done with a rather extensive set of compounds, including substituent effects, and hence is probably reasonably accurate [19]. However, one should remember that there is still some uncertainty in the parameters of any such calibration. For the heteronuclear  $^1\text{H}$ - $^{31}\text{P}$  and  $^1\text{H}$ - $^{13}\text{C}$  couplings there have also been relationships developed. These are less extensively calibrated, but are probably still sufficient to distinguish fairly well between *gauche* and *trans* conformers. Now that labelling with  $^{13}\text{C}$  is possible, there may be new  $^{13}\text{C}$ - $^{13}\text{C}$  or  $^{13}\text{C}$ - $^{31}\text{P}$  relationships derived which can also provide useful angle information.

The interpretation of NOEs to give distances is fairly straightforward in principle, but there are a number of practical concerns which make it difficult to do with the accuracy which is often desired. When there is an isolated pair of spins, and the tumbling of the molecule is slow ( $\omega\tau_c \gg 1$  where  $\omega$  is the Larmor frequency in angular units and  $\tau_c$  is the tumbling correlation time), the theory is quite simple. The NOE then initially (at short mixing times) builds up at a rate which is proportional to  $\tau_c/r^6$ . The measurement of a NOESY crosspeak volume then provides a slope, giving this rate. However there are still two unknowns,  $\tau_c$  and  $r$ . One can be eliminated by comparing a crosspeak between protons at a known distance (i.e. the H5 and H6 of cytosine or the H2' and H2'' of deoxyribose) with that of an unknown distance. Thus  $I_{\text{NOE1}}/I_{\text{NOE2}} = r_2^6/r_1^6$ .

This general approach is usually called the isolated spin pair approximation (ISPA), and it has two problems. First, it is explicitly assumed that each vector between a pair of spins tumbles isotropically, with the same correlation time. This is equivalent to saying that the molecule is spherical (not a good assumption for long, duplex oligomers [20]), and that there is no internal motion. In principle it is possible to correct for nonspherical shape through the introduction of two correlation times, which could be fit during the structure determination process. It is necessary to fit, as part of the calculation, the angle which each internuclear vector makes with the long axis of the molecule. This is an added computational complication, and has not been done in practice. The question of internal dynamics is complex. It has been shown in

proteins that fast internal vibration-like motions do not seriously affect the relaxation dynamics if the motions are reasonably isotropic and uncorrelated [21]. This is probably true to a good approximation for nucleic acids as well. However there may be other, larger amplitude motions occurring in nucleic acids which must also be considered. This question has been addressed through NOE measurements [22] (though arguments both for and against there being motion often come from almost the same data with different interpretation), through  $^{13}\text{C}$  and  $^{31}\text{P}$  relaxation measurements, and through  $^2\text{H}$  solid state NMR. The bulk of the data seem to suggest that there are usually no large amplitude motions of either the bases or the sugar which would seriously affect the NOE calculations.

The second assumption implicitly made in ISPA is that there are no other spins present to affect the cross relaxation. This is almost never the case, once an NOE develops between one pair of protons it can be subsequently passed to yet another through the same mechanism — a process known as spin diffusion. In a mathematical sense one can go to the short mixing time limit, in which linear NOE build-up should be observed, even if other spins are present. However in practice at sufficiently short mixing times the S/N of cross-peaks becomes very poor, and distances (particularly longer ones) cannot be determined very accurately. Again theory can come to the rescue, the full behavior of the relaxation can be calculated. If the structure is known, so that all interproton distances can be calculated, then it is possible to write the crosspeak amplitudes (in matrix form) as  $\mathbf{A} = 1 - e^{\mathbf{R}\tau}$ , where  $\mathbf{A}$  is the matrix of cross peak amplitudes,  $\tau$  is the NOE mixing time, and  $\mathbf{R}$  is the relaxation matrix. In principle, if  $\mathbf{A}$  is known completely and accurately, this equation can be inverted to give  $\mathbf{R}$ , and from it the distances (the unknowns of interest in a structure calculation) can be determined. Of course in practice the knowledge of  $\mathbf{A}$  is limited by S/N, overlap of cross and diagonal peaks, etc. so that the inversion cannot be perfectly done.

There have been a couple of algorithms developed which do solve this problem through an iterative procedure in which a guess is made at the structure (either from structure calculations with rough data, or from a model such as classical B DNA), which is then modified to give the best agreement with the experimental data [23]. Even within this approach there are two modifications. In one case the NOE amplitudes are fitted to give distances as accurately as possible (having included the effects of spin diffusion), which are then applied as constraints in a structure

calculation. During refinement of the structure reference is made back to the determined distances to calculate the error. In the alternate approach a comparison is made directly with the spectrum itself, optimizing the structure (a set of coordinates) to give the best fit to the NOE spectra. This is more analogous to what is done in X-ray diffraction where the electron density is used to generate a diffraction pattern, which is compared with the real one. An X-ray derived structure is considered "refined" when the difference between real and calculated diffraction patterns is minimized. The measure of the agreement is called the R-factor, which is the sum of the absolute values of the differences between real and calculated diffraction intensities divided by the sum of the intensities. For macromolecules a crystallographic R value in the neighborhood of 0.20 (or 20%) is accepted as indicating a reliable structure. For NMR there is another consideration, that is whether the NOE intensities should be used or the sixth root of the intensities. The problem with the intensities themselves is that short distances give very strong NOE peaks, which then dominate the error function. The sixth root weighting allows the weaker peaks to still have reasonable influence on the error function. Although there is not a formal "standard" yet, the latter approach seems to have the greatest following. It has recently been pointed out that intensity based "NMR R factor" has a hidden bias through the way it is defined, and an RMS form of the equation has been suggested [24].

In all of the discussion above it has been assumed that the NMR data have been carefully collected, processed and analyzed. During collection of data care should be taken that the spins are nearly fully relaxed between experiments, i.e. that the recycle delay is a few  $T_1$ s. For most protons of DNA this is not a particular problem, the  $T_1$  values are fairly uniform and short, excepting the adenosine H2s which are more isolated and have longer  $T_1$  values [25]. For short RNA oligomers the  $T_1$  values seem to be generally longer than for DNA, which could lead to errors in distance estimates if ignored [26]. Further considerations are whether the digital resolution is sufficient for proper representation of peak volumes, whether the apodization functions used affect the volume integrals nonuniformly, whether baseline artifacts affect integrations, and whether chemical exchange affects resonances. In general, *with proper caution*, none of these potential problems should have any major negative effect on the final structures determined.

#### 4. DEVELOPING MODELS

Given distances determined from NOEs and angles determined through Karplus relations, there are a number of ways to try to reconstruct a model (a set of coordinates) which agrees with the experimental data. The first method used in this task was metric space distance geometry. In this approach an algorithm is applied to convert interproton distances into distances to a "center of mass", then finally to a set of coordinates. When the distance information used is incomplete and uncertain (as is always the case with real data), then the initial coordinates do not satisfy the original input distance constraints. To bring them into agreement a refinement step is done, minimizing an error function which contains the experimental distance information, as well as information about the known covalent structure (fixed bond distances and angles). The concern with minimizing the error function is in finding the global minimum. This is now generally done with a simulated annealing approach, in which random velocities are assigned to atoms and Newton-like equations of motion are followed. After letting the atoms move for some time, the velocities are damped out and the agreement of the structure with input distances is checked. The process is repeated until the agreement between model and data is acceptable. There was some concern that the initial sampling of structures in the initial distance geometry phase was not as good as it could have been [27], however if the minimization phase is efficient many structures can be calculated, sampling widely in the space of conformations which agrees with the experimental data.

A more common version of this process now is to use a general molecular mechanics program, which has force fields describing the interactions of the constituent atoms of a macromolecule. In principle the energy calculated by such a program can be minimized (again there are major concerns of whether one can ever find the global minimum rather than local minima), to give "the structure" of the macromolecule. Unfortunately the potential energy functions are not really good enough to do this (although some believe that this will soon not be the case [28]), and some supplementary experimental information is needed. The distances determined from NMR can be included in the form of new "pseudo-potential" terms, where disagreement between the model and data increases the "energy". The search for structures agreeing with the NMR data, and at a minimum potential energy usually proceeds with molecular dynamics (much as the simulated annealing described above),

although an iterative minimization procedure has also been described [29]. A process for carrying out such a refinement using helical parameters [30] rather than conventional x,y,z coordinates has also been described, as well as a version which uses torsional angles [31]. This work has also examined the degree to which local structure variation might be detectable through NMR experiments. There have been efforts to examine the effects of local motion on the structure calculation [32]. The application of time average restraints for DNA has also been investigated [33]. It was found that relatively long MD trajectories were needed to properly characterize the dynamical averaging. In general it has been found that limited amplitude local motion causes relatively minor alterations in apparent distances, and hence probably does not introduce major uncertainty into structure calculations.

With any method used for generating structures, there are generally a family of models which agree essentially equally well with the data, and hence are all acceptable solutions. If real molecular mechanics programs are used then these structures are also near an energy minimum. The results of structure calculations are usually shown as a superposition of a set of these structures, overlapped for minimum rms deviation in their coordinates. A set of very closely grouped structures implies a high precision in the determination, while a broad spread indicates a less well defined structure. In interpreting these structures one must be careful to look at the source of the information used (distances only, with energy, number of NOEs? etc.) and how the calculations were done (minimization only, dynamics, artificial constraints? etc.).

The issue of internal dynamics of a nucleic acid was raised above, and somewhat swept aside with the statement that fast "vibration-like" motions probably do not affect NOE build-up very much and other motions probably do not occur very often. It is worth noting that if molecular dynamics calculations are properly done, which requires inclusion of explicit solvent and ions, then it is (in principle) possible to calculate NOEs with the correct inclusion of all dynamics. The predicted NOEs, calculated using a long molecular dynamics trajectory, could be compared with those observed experimentally. This approach would provide a correct, dynamic picture of the molecule. It has not been possible yet (basically due to computational limitations) to use the agreement with NMR during the calculation, although there are some programs which are approaching this. A remaining challenge will be to see that the potentials are good enough that they not only predict correctly where energy minima lie in conformational space, but also

that they correctly represent the barriers between minima, which are necessary to predict the kinetics correctly.

A final question related to the determination of nucleic acid structure in solution is: What quantity and quality of NMR determined restraints are necessary to determine the structure to the desired accuracy. Many more structure determinations have been done for proteins than for nucleic acids. For proteins it has been convincingly shown that a large number of rather loose constraints on interproton distance is sufficient to determine the structure of the protein rather accurately. As the number of the constraints increases the clustering of the structures improves. Improvement in the quality of the constraints also leads to an improvement in the quality of the structure. In nucleic acids the structural questions being posed are generally more than whether the molecule is in an "A" type or "B" type family. The nature of duplex nucleic acid structures is also different from proteins, there are neither tertiary interactions or a closely packed hydrophobic core which help constrain the geometry of the backbone. This makes it relatively more difficult to determine nucleic acid structures to an accuracy which provides structural features of real interest. It has been shown through test calculations that the set of distances used in many of the early structure analyses (involving aromatic protons and sugar H1', H2', H2'', H3') are insufficient to accurately reproduce the structure even if the distances are provided with extremely high accuracy [34] (in the absence of a potential energy function, although a hard sphere van der Waals "repulsion" is maintained). This means that structure calculations using such sets of distances in real systems must yield structures with a large degree of variation in many of the local conformational variables [35]. If the structures are tightly clustered this is an indication that there is a systematic bias in the structure calculation, and caution should be used in interpreting the results. If the calculations include energy minimization at the same time as constraint minimization, then there is better clustering of the structures [36]. The fact that test calculations return model structures (which were near energy minima to begin with) with good accuracy indicates an internal consistency. However, one must keep in mind that unusual structural features could be suppressed in such calculations due to the influence of the force field. In addition it is often necessary to manipulate the potential function somewhat (at least when calculations are done *in vacuo*) to avoid peculiar features. Such manipulations include "Charges on the 3' and 5' protons were modified to prevent unwanted bond

formation [36]”, or “we have added a few ‘generous’ distance restraints between C1’ atom on opposite sides of the major groove” because “the major groove had a tendency to collapse [37]”. These comments are not intended to imply that such manipulations should not be done, just that there are sufficient problems with the potential functions that work is required to obtain reasonable structures. One should then consider the extent to which the structures are affected by the NMR data, vs. the potential function. This was explicitly addressed in one set of calculations, in which it was shown that the NMR data do very significantly affect the structure obtained [38].

More recently it has been realized that some of the H4’, H5’ and H5’’ resonances can be assigned and either NOEs or J values can be used to generate constraints [39]. These can be “negative” constraints — if no NOE is observed at a known position then the separation between the appropriate protons must be more than ca. 4 Å. Such constraints remove some of the conformational variation in the phosphate positions, which in turn reduces the conformational variability elsewhere. The extent of improvement through use of such constraints has not yet been systematically studied in test calculations, but their inclusion can only improve the structures.

Much of the discussion above has been aimed at methods for the quantitative determination of duplex structures, certainly one much sought after goal. However it is worth noting that there are some nucleic acid structures which have features of real interest which can be observed even in “low resolution” structures. This includes the pattern of stacking and positions of bases in hairpin loop structures of both DNA and RNA, the patterns of interactions seen in complex RNAs such as the pseudoknot, whether mismatched or modified bases are stacked into or out of a duplex, and the patterns of stacking of bases in G-quartets. In all of these cases the basic sources of information are the same as described above. However since even the rough folding pattern is not known through examination of the sequence, even when coordinates in a model are imprecise features of interest can easily and reliably be determined. Examples are given below.

## 5. ANALYSIS OF DYNAMICS

The methods for detecting dynamic processes in nucleic acids are much the same as those used for proteins. The type of dynamic process and the rates which can be detected vary widely depending on the

approach chosen. The fastest processes are generally detectable with relaxation methods. Since protons suffer from spin diffusion, heteronuclear relaxation ( $^{13}\text{C}$ ,  $^{15}\text{N}$  or  $^{31}\text{P}$ ) is generally used. For  $^{13}\text{C}$  and  $^{15}\text{N}$  which have directly bonded protons, the heteronuclear dipolar coupling is the primary relaxation mechanism, which is well understood. By measuring  $T_1$ ,  $T_2$  and the heteronuclear NOE it is possible to describe motions in terms of an order parameter and an effective correlation time in a model free approach [40]. Since few nucleic acids have been isotopically enriched, and natural abundance measurements are time consuming and less precise, there are not yet detailed applications of this method in nucleic acids. However, due to new labelling methods, it can be expected that this will soon change, for RNA at least.

Determinations of hydrogen exchange rates (primarily for the imino protons) can provide information about fast, transient fluctuations in the structure. When the chemical events of this process are properly evaluated (base catalysis of the exchange event), then "opening" rates of the base pairs can be determined [41]. These rates are both sequence and structure dependent. Poly-A stretches, for example, have much slower opening rates than other A-T rich sequences [42]. Similarly some of the guanosines in G-quartets have very long exchange half-lives [43]. Such differences reflect differences in the events leading to the "open" state, from which exchange actually occurs. Unfortunately we have as yet a rather poor understanding of what this "open" state actually is, and hence the exchange is difficult to interpret in any quantitative way, although it is very valuable as a empirical marker in any case.

The chemical shift is a sensitive probe of chemical environment. When spins are transferred between different chemical environments, at a rate comparable to the difference in the frequencies in these environments, then chemical exchange broadening occurs. In cases where the frequency in each environment can be determined, it is possible to make an accurate determination of the rate of exchange. Such broadening has been observed in nucleic acids, in some cases representing an exchange between the duplex and single strand forms [44], and in other cases between two different duplex conformations [45]. While observation of such broadening demonstrates a dynamic process is present, the chemical shift cannot be interpreted quantitatively to allow a determination of what the "states" or conformers are which are exchanging. Such information must be developed through other observations, such as NOEs and coupling constants. Of course



these too will be averaged. For couplings the average is simply a population weighted average of the coupling constants in the conformers which are exchanging. For NOEs, when the process is slow compared with Larmor frequency, the average is over the build-up rate of the NOE, proportional to the inverse sixth power of the distance. If the rate is comparable with the Larmor frequency then the averaging is more complex, and the rate of motion in addition to the distances comes into play.

Conformational averaging of sugar residues can be detected through the determination of coupling constants. Since the sugars are cyclic structures, few variables are necessary to describe the conformation. Generally the two parameters used are a pseudorotation angle (roughly which atom is out of the plane) and amplitude (how far out of the plane the atom is). With these two parameters there are several different couplings which can be predicted using Karplus equations. If the measured couplings do not agree with any single conformation, then conformational averaging must be occurring. There may be a number of different populations and conformations which could give the same average values, but if one includes a little information about energetics, and assumes the smallest number of conformers giving the proper average is correct, then good models can be developed. Such conformational averaging is clearly seen in single strand oligomers, and often for the terminal bases in duplexes as well. Accurate coupling constant values are required to distinguish clearly between single conformers intermediate between the commonly observed C2' *endo* and C3' *endo* sugars seen in B-form and A-form oligomers respectively [46], and averaging between two sugar conformers.

A model has been developed in which there is a correlation between the phosphorous chemical shift and the torsional angles involving the phosphate [47]. Observation of differences in the chemical shifts of DNA phosphates in the *lac* repressor-operator complex has led to the suggestion that there are important dynamic differences between the proper sequence-specific complex, and others such as mutant operators. These dynamic differences are thought to contribute to the free energy of binding primarily in an entropic way [48]. Although perhaps still a controversial suggestion, NMR is one of the few methods which can detect site specific differences in structure and dynamics in a way which might lead to new insight into mechanisms for achieving sequence specificity in binding.

A rather different approach to determination of dynamics is the analysis of lineshapes of  $^2\text{H}$  in solid samples. In this case the quadrupole

coupling provides a line width and shape which are known (or can be determined) for any particular chemical site. When motion occurs this lineshape is averaged in a manner characteristic of the rate, amplitude and direction of the motion (although all three parameters cannot always be unambiguously determined). The linewidth is most sensitive to motions on the microsecond timescale, a range difficult to detect in solution. Attention must be taken to separate local and global motions, but comparison of labels at different locations make this possible in some cases. With this approach it has been possible to analyze changes in dynamics at different levels of hydration [11].

## 6. EXAMPLES OF STRUCTURE DETERMINATIONS

There have been many structures based upon NMR data published over the last few years. For some of these the assumptions made at the time the structure was done (ISPA, multiple correlation times, etc.) were reasonable, but have since been shown to introduce significant errors into distances, and hence structures. Some of the early structures were calculated just using distance geometry, but most now use restrained mechanics. The lack of standardization in both methods for calculating structures, and for reporting the agreement with the original data make many of these early structures difficult to evaluate in any quantitative way. For this reason the examples described here are all of rather recent vintage.

In spite of the increasing emphasis on quantitative structure determination, it is worth noting again that insight into features of nucleic acid structure can also come from a less rigorous analysis. Chuprina et al. [49] examined cross strand adenosine H2–sugar H1' distances in many DNA sequences, particularly those rich in A–T base pairs. They found that there was a good correlation between short values of this distance, and a narrow minor groove, albeit with considerable scatter in the data. The combined observations through both NMR and crystallography lead to suggestions for the relationship between sequence and groove width. In another study, a detailed comparison of the DNA oligomer (GCATTAATGC)<sub>2</sub>, and the equivalent sequence in which each adenosine was replaced with 2-amino-adenosine, showed that there was very little structural change between the parent and modified oligomers, in spite of the extra hydrogen bond in each base pair [50]. Mixed DNA/RNA sequences have also been studied by NMR [51],

including d(CGCG)r(AAUU)d(CGCG). In this case both the base stacking and the conformations of the sugars could be determined. Since the DNA segments prefer a C2' *endo* "B-form" conformation and the RNA a C3' *endo* "A-form" the conformation in the junction region is not obvious. From the analysis of the coupling constants it appears that the first ribonucleotide on each strand are perturbed from C3' *endo*. The other nucleotides behave much as they would be in either DNA or RNA homopolymers.

Hairpin loops (also known as stem-loops, a duplex in which strands are linked through any number of bases at one end) are an important structural element in higher order nucleic acid structure. This is particularly true for RNA, which is normally a single covalent strand which folds back upon itself in order to form all secondary structure. There have been a number of studies examining the relationship between loop length and overall hairpin stability. From the early studies it appeared moderate size loops (6–7 bases) seemed to be most stable, however structural data were not available to correlate particular structural features with stability. Since there is relatively little known about the positions of the bases in the loop regions, even a semiquantitative NMR analysis can provide important insight into these structures and why they form. There have now been a number of hairpin loops analyzed by NMR, including both DNA and RNA systems. In studies of DNA it was found that some loops, although nominally having four bases, actually contain only two bases which are not paired. For a loop containing TTTA, the base pair terminating the duplex forms with a T–A Hoogsteen pairing [52]. The Hoogsteen pairing is driven by the need to reduce the distance between the ends of the duplex so that it can be spanned by two nucleotides. The phosphate between the third and fourth nucleotides in the loop takes a *gauche* conformation for the  $\epsilon$  bond, as shown quite clearly through analysis of the coupling constants. The unusual geometry of the loop also leads to NOEs that are not seen in duplex regions of DNA. When a careful analysis was done of all of the NOEs and coupling constants it was found that the loop geometry could be defined quite well [53], including the fact that the T at the second position in the loop was folded back into the groove. Comparative studies showed that when the loop was ATTT rather than TTTA (reversal of the base pair closing the loop), that the Hoogsteen pairing was lost and a true four base loop was formed. With a loop sequence of CTTG, however, a two base loop did form, but with Watson–Crick rather than Hoogsteen pairing [52,54]. These studies of loops have lead

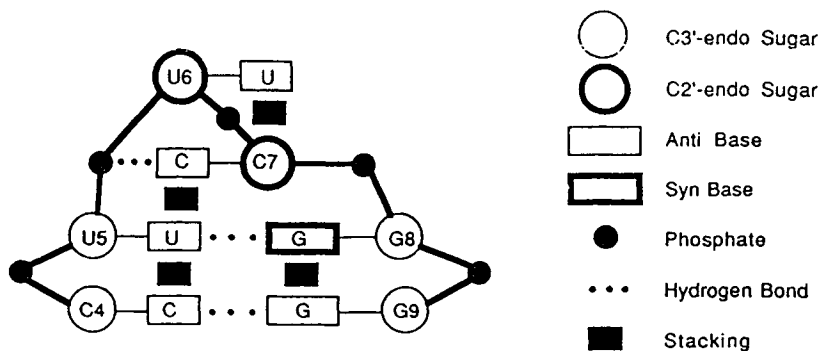


Fig. 6. A schematic structure is shown of a very stable RNA tetraloop, UUCG. The conformational features which were identified from the NMR data are indicated symbolically on the structure. From Varani et al. [56].

to some “rules” for loop formation and structure, but there remains much to learn about the details of other loops. The NMR methodology is clearly sufficient to provide structural information which greatly improves our understanding of these structures.

Some RNA tetraloops have been shown to occur frequently in natural RNAs, and it has been found that they are exceptionally thermostable [55]. When the UUCG loop was examined by NMR there were a number of structural features identified which provide insight into the basis for the stability. A G–U basepair forms in a reverse wobble conformation, giving two hydrogen bonds although there seems to be a relatively large buckle in the base pair. In addition to the pairing of the terminal bases of the stem, there is evidence for a hydrogen bond between the cytosine in the loop (one of the bases required for the unusual thermostability) and one of the phosphates. These were evident even from a semiquantitative analysis, and a distance geometry/energy minimization approach [56]. More recently this structure has been further refined [57], yielding further details of the structure. As for the DNA loops it is found that there is an unusual phosphate conformation, required to span the end of the loop. The two nucleotides in the center of the loop take on a C2'endo conformation, rather than the usual C3'endo seen in RNA. This leads to a longer distance between the attached phosphates, again making it possible for two residues to reach across the stem. Another of the frequently occurring tetraloops, GNRA (where N is any base and R is a purine) has also been examined by NMR [58]. The results show that there is a G–A base pair, in which there are hydrogen bonds from the G

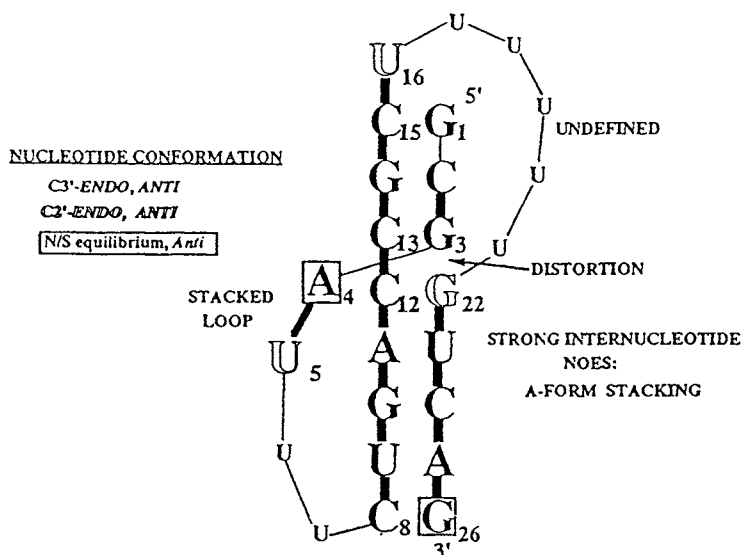


Fig. 7. A schematic structure is shown of an RNA pseudoknot. The features of the structure identified through NMR are indicated [59].

amino proton to the A N7, and from the A amino to the G N3. In addition there is a hydrogen bond from the G to a phosphate in the loop, and between the 2'OH of the sugar and the A across the loop. These data explain much about the roles of the conserved bases, and the interactions stabilizing the loops.

The general methodology has been applied to another RNA structure, the pseudoknot [59]. These are formed by sequences which form a hairpin loop, but have a tail which folds back to interact with the loop regions (Fig. 7). Although the larger size of this structure made the assignments and structure analysis somewhat more difficult, most of the aromatic and sugar H1' resonances could be assigned, making it possible to analyze both sugar conformation and stacking. It was found that the two stem regions costack to make an almost continuous helix, much as the acceptor and Ty stems of tRNA costack. In order to accommodate the stacking interactions, and to span the ends of the stems with the residues available, there are again sugars which convert to the C2'*endo* conformation. Interactive graphics with a molecular mechanics program were used to generate a model which agreed with the NMR observations. Although this is not a high resolution structure,

it does provide important insight into the nature of the structure, and suggests further experiments to relate this type of structure to functional properties.

Work has begun on other larger RNA structures as well. 5S ribosomal RNA has been examined by several different groups. For many years obtaining assignments was the limiting factor, however with high field spectrometers, labelling and more facile RNA synthesis the assignments have been obtained. The helix I fragment of *E. coli* 5S RNA has been analyzed in detail, applying NOEs (with relaxation matrix analysis), and coupling constants, with the XPLOR molecular dynamics program [60]. The structure is shown to be an A-form helix, as anticipated, although there is deviation from the regular helix geometry on one side of a G–U wobble base pair. Energy minimization and molecular dynamics (using BIOGRAF) have been applied to helix II–III region, and helix IV regions of *B. megaterium* 5S RNA as well [61]. Although not yet a high resolution analysis, the NMR data are allowing structural features to be identified. Self-cleaving “hammerhead” RNAs have also been analyzed using NMR by several groups [62]. At this stage only features of the secondary structure, and hints at dynamics and tertiary structure have been discerned. However with labelling methodology, and 3D or 4D NMR it should be possible to analyze these larger molecules as well.

Some of the early duplex DNA structures determined with NMR suffered from incomplete understanding of some of the problems discussed in the methodology sections above. For this reason, although the convergence described in the papers seems reasonable, the structural features are probably not really accurately determined. The introduction of more extensive analysis for distance determinations (multiple mixing times, relaxation matrix analysis, and inclusion of angles), and more sophisticated refinement methods (restrained MD), seems to be leading to structures which are more reliable. It should be remembered in these that there is a significant effect of the potential function in the MD calculations, and hence these structures are really not fully determined from the NMR data although they are clearly heavily influenced by it. James' group has carried out structure determinations of a number of duplex oligomers, including d(ACACACAC):d(GTGTGTGT) [63], d(GTATATAC)<sub>2</sub> [64], d(ATATATAUAT)<sub>2</sub> [65], d(GTATAATG):d(CATATTAC) [66]. With the extensive NMR characterization and analysis now done it seems possible to discern local features of interest. There are significant deviations from the average or canonical B-DNA structure, and

these may be important for the recognition of particular DNA sequences by proteins, or in their ability to interconvert between different conformations in solution. In another detailed study of an octamer DNA, d(GCGTTTCGC):d(CGCAACGC), the effects of dynamics on the structure calculation were included [32]. As noted in the discussion of dynamics this did lead to a somewhat better agreement between the real and simulated spectrum, but did not make major difference in the resulting structure.

The structural features of a number of DNAs containing mismatched base pairs, extra bases on one strand, or a chemically modified base, have been determined with NMR. This area has been particularly profitable in that there are relatively few crystal structures of such molecules, but they play important roles in mutagenesis. In addition, in a sequence containing an extra adenosine there have been significant differences between the NMR solution structures [67] (the extra A stacked into the helix) and crystal structure [68] (the extra A out of the helix). Other recent examples are the analysis of a double G<sup>+</sup>A mismatch [69], and an oligomer with a derivatized guanosine opposite adenosine [70]. These systems are again ones in which even semiquantitative structures can lead to important conclusions about the biological behavior of the molecules.

Multiple strand DNA structures have also been examined in some detail. The first of these were triple strands, in which one polypurine strand can pair with two pyrimidine strands, giving T:A–T and C<sup>+</sup>:G–C base triples. Early NMR studies used samples in which the pyrimidine strand was added two-fold over the purine strand. However a later construction used a single strand, which folds back on itself to give the triple strand with loops at each end. Although the general pairing of the bases in such triple strands had been established long ago from fiber diffraction data, the details of the local structure, particularly the sugar-phosphate backbone, were not known well. The NMR data have provided a much clearer picture of these structures. When DNA sequences found in telomeres (the extreme ends of chromosomes, generally rich in G and T, and not Watson–Crick base paired) were studied as oligomers, interesting aggregation phenomena were seen. In native gels the telomeric sequences had high mobility which indicated a folded structure, and that the stability of the structures were cation dependent. Initial NMR studies indicated that there were guanosine bases in both *anti* and *syn* glycosidic conformations [71]. Further work led to the idea that the guanosines were forming an intramolecular tetraplex

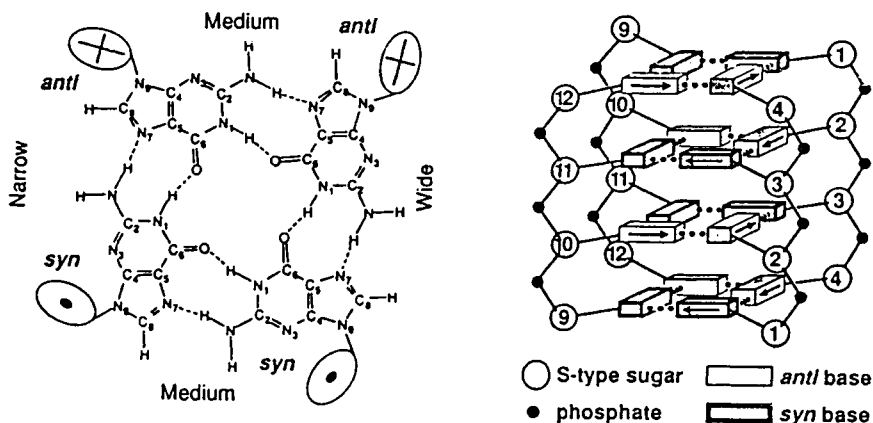


Fig. 8. The structure of a telomeric DNA sequence "G-quartet" tetraplex structure is shown. Experimental constraints were derived from NOE and J-coupling data.

structure, again something first characterized at low resolution in fibers using x-ray diffraction. Very recently a structure based upon NMR has been published (Fig. 8) giving a higher resolution view of this tetraplex [72]. Interestingly there was a crystal structure published in the same journal issue, which is qualitatively similar, but different in a number of respects as well. It may again be the case that crystal packing forces are sufficient to introduce differences between the solution and crystal structures.

Structure determinations of nucleic acids in complex with drugs and proteins have also advanced dramatically in the past few years. The antibiotic chromomycin was shown to be a dimer when bound to DNA [73], the two equivalent molecules associated through a bound magnesium ion. NOE contacts between the chromomycins and DNA showed clearly that the drug was bound in the minor groove, rather than the major groove as initially believed. A detailed model for this complex was then developed [74], which showed that there are significant distortions of the DNA induced by the drug binding, particularly a widening of the minor groove in the binding site. Contacts between the chromomycin molecules and the DNA can be seen, which explain several features of the sequence specificity in binding.

Extensive NMR studies have been carried out on other minor groove binding compounds as well. Titration studies with distamycin-A showed that there were several different types of complex formed when it bound



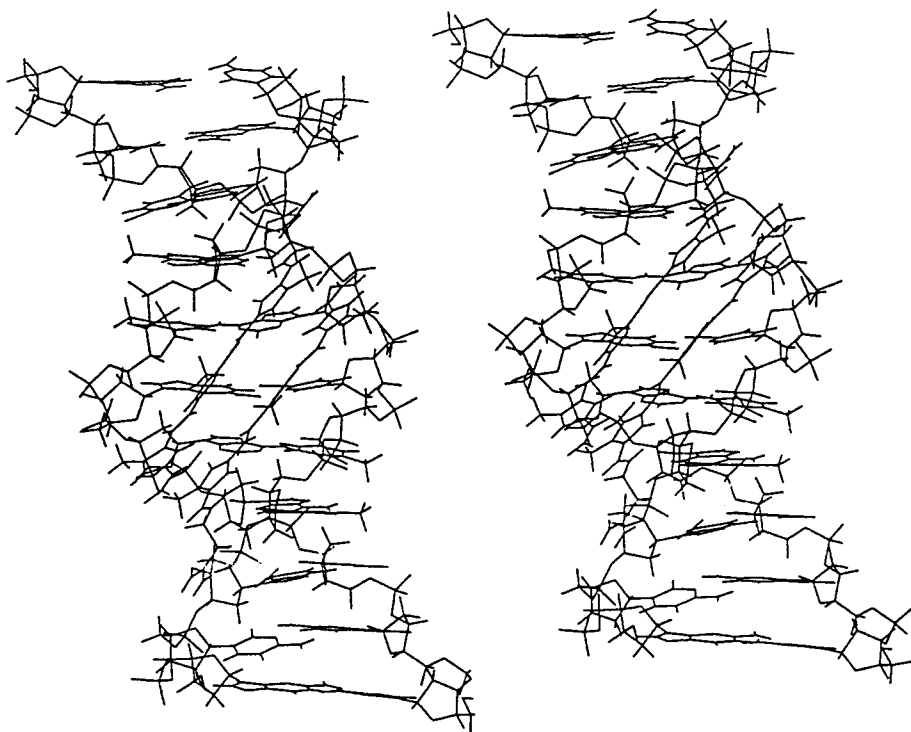


Fig. 9. The structure of a G-C binding distamycin analog-DNA complex is shown. The structure was determined using NMR derived distance constraints, and molecular modelling [77].

to an oligomer containing a binding site AAATT:AATTT. It was shown that at low drug/DNA ratios ( $< 0.5$ ) the drug binds in two orientations with one distamycin per DNA [75], and exchange of the distamycin occurs between neighboring subsites of four ATs. At higher amounts of added distamycin detailed NOE studies showed that a complex with two drug molecules side-by-side in the binding site can be formed [76]. Studies of a number of other sequences lead to the idea that groove width affects the relative affinity of the 1:1 and 2:1 modes. When this idea was applied to modified distamycin analogs, with a pyrrole ring replaced by an imidazole, G:C base pairs at specific sites could be recognized [77]. In addition to these non-covalent complexes there have been numerous covalently modified DNAs studied, including adducts of tomaymycin [78], CC-1065

[79], and photomodified DNA containing thymine dimers [80] and psoralen crosslinks [81]. Although many of these have been at the low resolution level, further application of the more recently developed methodology will allow them to be determined in more detail.

Work with protein-DNA complexes has been less extensive, although there is rapid progress in a number of systems. There have been several complexes in which contacts between protein and DNA have been identified, and have allowed modelling of the complex. The *lac* repressor headpiece was one of the first DNA binding proteins studied, and its binding orientation with respect to the operator sequence was determined from the NMR data [82], coming out the opposite from what had been proposed based on biochemical data. The *antennapedia* homeodomain protein complex with DNA has also been studied. Again, although the number of contacts was not large, the binding orientation could be determined and a number contacts between specific sidechains and bases identified [83]. These studies are made difficult by the high complexity of both the protein and DNA, and have been helped tremendously by isotope filtering methods.

## 7. SUMMARY

Overall, although there are special problems with determination of nucleic acid structures to high accuracy, NMR is now a valuable tool for structural analysis of DNA and RNA, as well as their complexes with drugs and proteins. When care is exercised in the development of structural constraints from the NMR data, and as many restraints are determined as possible, many interesting features of the fold and interactions of the nucleic acids can reliably be determined. It is likely that with further implementation of isotope labelling methods, and 3D and 4D NMR methods both the molecular weight which can be analyzed, and the accuracy of the structures determined can be extended.

## REFERENCES

1. F.M.J. van de Ven and C.W. Hilbers, *Eur. J. Biochem.*, 178 (1988) 1.
2. J. Feigon, W. Leupin, W.A. Denny and D.R. Kearns, *Biochemistry*, 22 (1983) 5943; R.M. Scheek, N. Russo, R. Boelens and R. Kaptein, *J. Am. Chem. Soc.*, 105 (1983) 2914; D.R. Hare, D.E. Wemmer, S.-H. Chou, G. Drobny and B.R. Reid, *J. Mol. Biol.*, 171 (1983) 319; W.J. Chazin, K. Wüthrich, S. Hyberts, M. Rance, W.A. Denny and W. Leupin, *J. Mol. Biol.*, 190 (1986) 439.

3. K. Wüthrich, *NMR of Proteins and Nucleic Acids*. Wiley, New York, 1986.
4. H.A. Hues, A. Pardi, *J. Am. Chem. Soc.*, 113 (1991) 4360.
5. S. Chem, F. Heffron, W. Leupin, W.J. Chazin, *Biochemistry*, 30 (1991) 766.
6. V. Sklenár and J. Feigon, *J. Am. Chem. Soc.*, 112 (1990) 5644.
7. P.B. Moore, personal communication.
8. Radhakrishnan, D.J. Patel and X. Gao, *Biochemistry*, 31 (1992) 2514.
9. R. Boelens, G.W. Vuister, T.M.G. Koning and R. Kaptein, *J. Am. Chem. Soc.*, 111 (1989) 8525.
10. C.K. Brush, M.P. Stone and T.M. Harris, *Biochemistry*, 27 (1988) 115.
11. R. Brandes, R.R. Vold, R.L. Vold and D.R. Kearns, *Biochemistry*, 25 (1986) 7744; A. Kintanar, W.C. Huang, D.C. Schindele, D.E. Wemmer and G. Drobny, *Biochemistry*, 28 (1989) 282; T.M. Alam and G. Drobny, *Biochemistry*, 29 (1990) 3421.
12. G. Varani and I. Tinoco, Jr., *J. Am. Chem. Soc.*, 113 (1991) 9349.
13. J. Ashcroft, S.R. LaPlante, P.N. Borer and D. Cowburn, *J. Am. Chem. Soc.*, 111 (1989) 363.
14. E.R. Kellenbach, M.L. Remerowski, D. Eib, R. Boelens, G.A. van der Marel, H. van den Elst, J.H. van Boom and R. Kaptein, *Nuc. Acids Res.*, 20 (1992) 653.
15. R.H. Griffey, C.D. Poulter, Z. Yamaizumi, S. Nishimura, R.E. Hurd, *J. Am. Chem. Soc.*, 104 (1982) 5810.
16. G. Otting and K. Wüthrich, *Quart. Rev. Biophys.*, 23 (1990) 39.
17. E.P. Nikonowicz and A. Pardi, *Nature*, 355 (1992) 184.
18. K. Gehring, P.G. Williams, J.G. Pelton, H. Morimoto and D.E. Wemmer, *Biochemistry*, 30 (1991) 5524.
19. C.A.G. Haasnoot, F.A.A.M. de Leeuw and C. Altona, *Tetrahedron*, 36 (1980) 2783.
20. A.J. Birchall and A.N. Lane, *Eur. J. Biophys.*, 19 (1990) 73.
21. D.M. LeMaster, L.E. Kay, A.T. Brünger and J.H. Prestegard, *FEBS Lett.*, 236 (1988) 71.
22. B.R. Reid, K. Banks, P. Flynn and W. Nerdal, *Biochemistry*, 28 (1989) 10001.
23. B.A. Borgias and T.L. James, *J. Magn. Reson.*, 79 (1990) 493; E.P. Nikonowicz, R.P. Meadows and D.G. Gorenstein, *Biochemistry*, 29 (1990) 4193; J.D. Baleja, J. Moult and B.D. Sykes, *J. Magn. Reson.*, 87 (1990) 375; R. Boelens, T.M.G. Koning and R. Kaptein, *J. Molec. Struct.*, 173 (1988) 299; R. Boelens, T.M.G. Koenig, G.A. van der Marel, J.H. van Boom and R. Kaptein, *J. Magn. Reson.*, 82 (1989) 290.
24. J.M. Withka, J. Srinivasan and P.H. Bolton, *J. Magn. Reson.*, 98 (1992) 611.
25. K. Weisz, R.H. Shafer, W. Egan and T.L. James, *Biochemistry*, 31 (1992) 7477.
26. A.C. Wang, S.G. Kim, P.F. Flynn, S.-H. Chou, J. Orban and B.R. Reid, *Biochemistry*, 31 (1992) 3940.

27. W.J. Metzler, D.R. Hare and A. Pardi, *Biochemistry*, 28 (1989) 7045.
28. J.M. Withka, S. Swaminathan, J. Srinivasan, D.L. Beveridge and P.H. Bolton, *Science*, 255 (1992) 597.
29. H. Robinson and A.H.-J. Wang, *Biochemistry*, 31 (1992) 3524.
30. N.B. Ulyanov, A.A. Gorin, V.B. Zhurkin, B.-C. Chen, M.H. Sarma and R.H. Sarma, *Biochemistry*, 31 (1992) 3918.
31. R. Ajay Kumar, R.V. Hosur and G. Govil, *J. Biol. NMR*, 1 (1991) 363.
32. T.M.G. Koning, R. Boelens, G.A. van der Marel, J.H. van Boom and R. Kaptein, *Biochemistry*, 30 (1991) 3787.
33. D.A. Pearlman and P.A. Kollman, *J. Mol. Biol.*, 220 (1991) 457.
34. D.E. Wemmer, unpublished.
35. A. Pardi, D.R. Hare and C. Wand, *Proc. Nat. Acad. Sci.*, 85 (1988) 8785; W.J. Metzler, C. Wand, D.B. Kitchen, R.M. Levy and A. Pardi, *J. Mol. Biol.*, 214 (1990) 711.
36. K. Kaluarachchi, R.P. Meadows and D.G. Gorenstein, *Biochemistry*, 30 (1990) 8785.
37. J.D. Baleja, R.T. Pon and B.D. Sykes, *Biochemistry*, 29 (1990) 4828.
38. A.M. Gronenborn and G.M. Clore, *Biochemistry*, 28 (1989) 5978.
39. S.-G. Kim, L.-J. Lin and B.R. Reid, *Biochemistry*, 31 (1992) 3564.
40. G. Lipari and A. Szabo, *J. Am. Chem. Soc.*, 104 (1982) 4546; G. Lipari and A. Szabo, *J. Am. Chem. Soc.*, 104 (1982) 4560.
41. J.L. Leroy, D. Broseta and M. Guéron, *J. Mol. Biol.*, 184 (1985) 165; J.L. Leroy, M. Kochoyan, T. Huynh-Dinh and M. Guéron, *J. Mol. Biol.*, 200 (1988) 223.
42. J.L. Leroy, E. Charretier, M. Kochoyan and M. Guéron, *Biochemistry*, 27 (1988) 8894.
43. F.W. Smith and J. Feigon, *Nature*, 356 (1992) 164.
44. J.M. Pieters, J.R. Mellema, H. van den Elst, G.A. van der Marel, J.H. van Boom and C. Altona, *Biopolymers*, 28 (1989) 717.
45. A.N. Lane, *Biochem. J.*, 259 (1989) 715; J.F. Lefevre, A.N. Lane and O. Jardetzky, *Biochemistry*, 27 (1988) 1086.
46. L.J. Rinkel and C. Altona, *J. Biomol. Struct. Dyn.*, 4 (1987) 621.
47. V.A. Roongta, R. Powers, E.P. Nikonowicz, C.R. Jones and D.G. Gorenstein, *Biochemistry*, 29 (1990) 5245.
48. C. Karslake, M.V. Botuyan and D.G. Gorenstein, *Biochemistry*, 31 (1992) 1849.
49. V.P. Chuprina, A.A. Lipanov, O.Y. Fedoroff, S.-G. Kim, A. Kintanar and B.R. Reid, *Proc. Natl. Acad. Sci. USA*, 88 (1991) 9087.
50. W.J. Chazin, M. Rance, A. Chollet and W. Leupin, *Nuc. Acids Res.*, 19 (1991) 5507.
51. S.-H. Chou, P. Flynn, A. Wang and B.R. Reid, *Biochemistry*, 30 (1991) 5248.
52. M.J.J. Blommers, J.A.L.I. Walters, C.A.G. Haasnoot, G.A. van der Marel,

- J.H. van Boom and C.W. Hilbers, *Biochemistry*, 28 (1989) 7491.
53. M.J.J. Blommers, F.J.M. van de Ven, G.A. van der Marel, J.H. van Boom and C.W. Hilbers, *Eur. J. Biochem.*, 201 (1991) 33.
  54. M. Amaratunga, E. Snowden-Ifft, D.E. Wemmer and A.S. Benight, *Biopolymers*, 32 (1992) 865.
  55. C. Tuerk, P. Gauss, C. Thermes, D.R. Groebe, M. Gayle, N. Guild, G. Stormo, Y. D'Aubenton-Carafa, O.C. Uhlenbeck, I. Tinoco, Jr., E.N. Brody and L. Gold, *Proc. Natl. Acad. Sci.*, 85 (1988) 1364.
  56. C. Cheong, G. Varani and I. Tinoco, Jr., *Nature*, 346 (1990) 680.
  57. G. Varani, C. Cheong and I. Tinoco, Jr., *Biochemistry*, 30 (1991) 3280.
  58. H.A. Hues and A. Pardi, *Science*, 253 (1991) 191.
  59. J.D. Puglisi, J.R. Wyatt and I. Tinoco, Jr., *J. Mol. Biol.*, 214 (1990) 437.
  60. S.A. White, M. Nilges, A. Huang, A.T. Brünger and P.B. Moore, *Biochemistry*, 31 (1992) 1610.
  61. J.H. Kim and A.G. Marshall, *Biopolymers*, 32 (1992) 1263.
  62. H.A. Heus, O.C. Uhlenbeck and A. Pardi, *Nuc. Acids Res.*, 18 (1990) 1103; A.C. Pease and D.E. Wemmer, *Biochemistry*, 29 (1990) 9039; O. Odai, H. Kodama, H. Hiroaki, T. Sakata, T. Tanaka and S. Uesugi, *Nuc. Acids Res.*, 18 (1990) 5955; H.A. Hues and A. Pardi, *J. Mol. Biol.*, 217 (1991) 113.
  63. M. Gochin and T.L. James, *Biochemistry*, 29 (1990) 11172; M. Gochin, G. Zon and T.L. James, *Biochemistry*, 29 (1990) 11161.
  64. U. Schmitz, D.A. Pearlman and T.L. James, *J. Mol. Biol.*, 221 (1991) 271; U. Schmitz, G. Zon and T.L. James, *Biochemistry*, 29 (1990) 2357.
  65. D.J. Kerwood, G. Zon and T.L. James, *Eur. J. Biochem.*, 197 (1991) 583.
  66. U. Schmitz, I. Sethson and T.L. James, *J. Mol. Biol.*, 227 (1992) 510.
  67. M.W. Kalnik, D.G. Norman, P.F. Swann and D.J. Patel, *J. Biol. Chem.*, 264 (1989) 3702.
  68. L. Joshua-Tor, D. Rabinovich, H. Hope, F. Frolow, E. Appella and J.L. Sussman, *Nature*, 334 (1988) 81.
  69. E.P. Nikonowicz, R.P. Meadows, P. Fagan and D.G. Gorenstein, *Biochemistry*, 30 (1991) 1323.
  70. P. Huang and M. Eisenberg, *Biochemistry*, 31 (1992) 6518.
  71. E. Henderson, C.C. Hardin, S.K. Wolk, I. Tinoco, Jr. and E.H. Blackburn, *Cell*, 51 (1987) 899.
  72. F.W. Smith and J. Feigon, *Nature*, 356 (1992) 164.
  73. X. Gao and D.J. Patel, *Biochemistry*, 28 (1989) 751.
  74. X. Gao, P. Mirau and D.J. Patel, *J. Mol. Biol.*, 223 (1992) 259.
  75. J.G. Pelton and D.E. Wemmer, *J. Biomol. Struct. Dyn.*, 8 (1990) 81.
  76. J.G. Pelton and D.E. Wemmer, *Proc. Natl. Acad. Sci. USA*, 86 (1989) 5723.
  77. P. Fagan and D.E. Wemmer, *J. Am. Chem. Soc.*, 114 (1992) 1080; T.J. Dwyer, B.H. Geierstanger, Y. Bathini, J.W. Lown and D.E. Wemmer, *J. Am. Chem. Soc.*, 114 (1992) 5911; M. Mrksich, W.S. Wade, T.J. Dwyer, B.H. Geierstanger, D.E. Wemmer and P.B. Dervan, *Proc. Natl. Acad. Sci.*,

- 89 (1992) 7586.
78. F.L. Boyd, D. Stewary, W.A. Remers, M.D. Barkley and L.H. Hurley, *Biochemistry*, 29 (1990) 2387.
  79. C.H. Lin, G.C. Hill and L.H. Hurley, *Chem. Res. Toxicol.*, 5 (1992) 167.
  80. J. Kemmink, R. Boelens, T.M.G. Koning, R. Kaptein, G.A. van der Marel and J.H. van Boom, *Eur. J. Biochem.*, 162 (1987) 37.
  81. M.T. Tomic, D.E. Wemmer and S.H. Kim, *Science*, 238 (1987) 1722; P. Spielmann, T.J. Dwyer, J.E. Hearst, D.E. Wemmer, unpublished.
  82. R.M. Lamerichs, R. Boelens, G.A. van der Marel, J.H. van Boom and R. Kaptein, *Eur. J. Biochem.*, 194 (1990) 629.
  83. G. Otting, Y.Q. Qian, M. Billeter, M. Muller, M. Affolter, W.J. Gehring and K. Wüthrich, *EMBO J.*, 9 (1990) 3085.

## Chapter 7

# Carbohydrate Structure and Dynamics from NMR

LAURA E. LERNER

### ABSTRACT

High resolution NMR spectroscopy is a powerful method for the analysis of carbohydrates. Methods for determination of composition, sequence, conformation, and dynamics of oligosaccharides and glycoconjugates are reviewed here. Recent innovations such as selective excitation, three-dimensional pulse sequences, and gradients are illustrated with numerous examples from the literature.

### 1. INTRODUCTION

The importance of carbohydrates in biology is gaining increasing recognition. Carbohydrates are involved in a diverse array of processes involving cell recognition, including cell adhesion; clearance of circulating glycoproteins; binding of antibodies, pathogen, toxins; lymphocyte homing; and differentiation. New structures occurring on glycoproteins and proteoglycans are being reported continually.

Why are oligosaccharides emerging as central players in biological recognition? Because of the number of linkages possible between hexopyranose rings, a handful of monosaccharide building blocks can be combined in many chemically distinct combinations, providing oligosaccharides with great capacity for specificity. One of the best known examples of this capacity is human blood types, which are determined by the specific antigenic oligosaccharides on blood cell surfaces.

As yet, there seem to be as many functions for oligosaccharides as there are types of oligosaccharides [1]. In some cases, changing or removing oligosaccharides from glycoproteins has no detectable effect on the function of the protein part; while in others, alteration of the glycosylation completely destroys function [1]. Despite the lack of a common function, there is no doubt that oligosaccharides are important in many disease processes. For example, it has been shown recently that HIV-1 infected T-lymphocytes produce several membrane proteins with altered glycosylation patterns [2]. There is evidence that the glycosyl side chains on glycoproteins of the envelope of HIV-1 can interfere with binding of neutralizing antibodies, which would provide the mechanism for impairment of the immunological response in HIV-infection [3]. HIV-1 and HIV-2 envelope glycoproteins have been shown to bind mannosyl residues, N-acetylglucosamine, and mannose-6-phosphate [4,5]. Since mannose-6-phosphate is a marker of lysosomal enzymes, it is possible that this binding could interfere with lysosomal enzyme sorting in HIV-infected cells [5].

There is an ever-increasing need for better analytical methods for determining the structure of oligosaccharides in solution: not just their primary structure, or sequence of sugar residues, but also their three-dimensional conformation and their dynamics. Understanding conformation and dynamics is essential for studying the interactions of oligosaccharide ligands with their receptors. Identification of glycosylation patterns of engineered proteins is critical for the biotechnology industry [6] because many proteins of therapeutic potential are actually glycoproteins. It is important to determine if expression of the protein alters its glycosylation and how such alteration may affect function.

The two major methods used for obtaining the primary structure (sequence) of oligosaccharides are mass spectrometry and NMR [7]. It is possible to use chemical degradation coupled with various chromatographic methods to identify composition: that is, how many glucosyl residues, how many galactosyl residues, etc. By chemical modification and mass spectrometry, it is possible to add information about the sequence of these residues. (For a recent review of methods for analyzing oligosaccharides from glycoproteins, see Ref. [8].) However, only NMR spectroscopy offers a non-destructive method for determining not only composition and sequence, but also anomeric configuration, conformation, and dynamics of oligosaccharides. The drawback of NMR is its relative insensitivity, requiring milligram quantities of pure, homogeneous material. As Serianni has pointed out [9], a non-destructive



method that uses 5–10 mg has little practical advantage over a destructive method that requires only micrograms.

In this chapter, NMR methods for obtaining all this information will be summarized. The intended audience is NMR spectroscopists who are not familiar with carbohydrates, and carbohydrate chemists who do not use NMR. The goal is to provide this audience with a starting point for entering the extensive literature on the application of NMR to carbohydrates. (Readers should also be aware of the series, *Carbohydrate Chemistry*, Chemical Society, London, which summarizes publications dealing with NMR of carbohydrates each year.) The principles underlying multi-dimensional methods have already been reviewed in previous chapters. In this chapter, methods that have proven to be especially useful for analysis of oligosaccharides will be described. Since several excellent reviews covering some of these topics are already available [9–15], the emphasis here will be on more recent developments. Also, the emphasis will be on analysis of complex oligosaccharides found in animals and bacteria, as opposed to polysaccharides; and on pyranoses (six-membered ring sugars) rather than furanoses (five-membered ring sugars) found in nucleic acids (see Chapter 6). The only practical consequence of this emphasis is the choice of examples—most methods described will apply regardless of the source of sample.

When confronted with an unknown oligosaccharide sample to analyze, the NMR spectroscopist's first task is assignment of  $^1\text{H}$  and  $^{13}\text{C}$  resonances so that the sugar residues can be identified. In some cases, the magnitudes of certain  $^3J_{\text{HH}}$  values are also used to identify particular residues. The next level of information is the location and anomeric configuration ( $\alpha$  or  $\beta$ ) of the interglycosidic linkages (sequence). The location of the linkages can be determined most directly via long-range  $^1\text{H}$ - $^{13}\text{C}$  correlation spectroscopy (HMBC; see below). The anomeric configuration can be determined by the  $^1\text{H}$  chemical shifts and  $^3J_{\text{H}_1\text{H}_2}$  coupling constants, by characteristic  $^{13}\text{C}$  chemical shifts and  $^1J_{\text{CH}}$  values characteristic of  $\alpha$  or  $\beta$  linkages, and/or by interglycosidic NOE or ROE intensities. Once the sequence and linkage types are known, it may be of interest to determine the three-dimensional shape (conformation) of the oligosaccharide. This would require determination of interglycosidic NOEs or ROEs, homonuclear and heteronuclear coupling constants, and  $^1\text{H}$  and  $^{13}\text{C}$  relaxation parameters, usually in conjunction with molecular mechanics or dynamics calculations.

## 2. ASSIGNMENT

Some monosaccharide residues typically found in mammalian oligosaccharides are shown in Fig. 1. Nomenclature for monosaccharides, oligosaccharides, and glycoconjugates was recently reviewed by Kisailus and Allen [16].

Although there may seem to be only a few basic building blocks, the potential for great specificity in oligosaccharides arises from the many variations in side group orientation and in potential linkage points on these pyranoses. Unfortunately for the purposes of analysis, such rather subtle changes in orientation lead to very little chemical shift dispersion

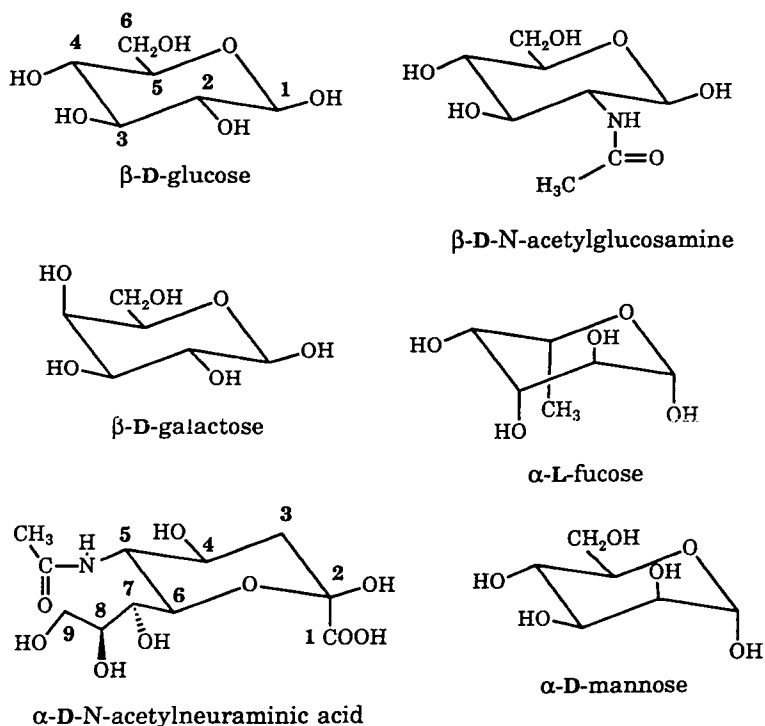


Fig. 1. Monosaccharides (pyranose forms) frequently found in mammalian oligosaccharides. Ring protons are omitted for clarity. Carbon positions are numbered for  $\beta$ -D-glucose and  $\alpha$ -D-N-acetylneuraminic acid. Commonly used abbreviations are: glc, glucose; gal, galactose; man, mannose; GlcNAc, N-acetylglucosamine; fuc, fucose; NANA or NeuAc, N-acetylneuraminic acid (sialic acid)

for ring protons; that is, the typical ring proton is bound to a carbon which is in turn bound to an hydroxyl group, so most ring protons have overlapping chemical shifts between 3 and 4 ppm. This makes assigning proton spectra for oligosaccharides much more challenging than for oligopeptides or oligonucleotides of comparable molecular weight.

### **2.1. One-dimensional methods**

Some carbohydrate protons do have distinctive chemical shifts. Early studies established that  $\alpha$ -anomeric protons resonate downfield of  $\beta$ -anomeric protons, and that both are well downfield of ring protons [15,17–20]. Another useful example is that the methylene protons on carbon-3 of sialic acid, a common constituent of cell-surface oligosaccharides, have distinct chemical shifts (ca. 1.8 ppm for the axial proton and 2.75 for the equatorial proton; see Fig. 5). These and other distinctive resonances can be used as “structural reporters” to identify residues within an oligosaccharide. Before the widespread use of multi-dimensional NMR, assignment of protons relied heavily on this structural reporter group approach, developed mainly by Vliegthart and co-workers [19,20]. Tables of characteristic  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts can be found in several reviews [15,19–22]. Chemical shifts of hydroxyl protons may also be useful in identifying residues [23]. Hydroxyl protons can be observed in solvents without exchangeable protons or in aqueous solvent mixtures by the application of appropriate solvent suppression pulse sequences and low temperature to slow exchange [24,25].

Simple comparison of one-dimensional  $^1\text{H}$  and/or  $^{13}\text{C}$  spectra can still be a useful approach when the oligosaccharide in question is expected to be identical with or to contain a known structure. Comparison of one-dimensional  $^1\text{H}$  spectra would be the simplest, fastest method of quality control for checking synthetic oligosaccharides, for detecting changes in post-translational modification of oligosaccharides obtained from glycoproteins, or for screening markers for clinical diagnoses. Several groups are developing automated assignment algorithms for application to oligosaccharides [26–31].

Selective deuteration can aid in assignment either directly [32–34] or by its effect on  $^1\text{H}$  or  $^{13}\text{C}$  relaxation [35–40]. Another use of deuteration for the purpose of assignment is the differential isotope shift (DIS) technique which relies on deuterium-induced  $^{13}\text{C}$  chemical shifts produced when deuterium exchanges with an hydroxyl proton. Pfeffer and

coworkers [41] reported DIS values for many pyranose and furanose mono- and disaccharides. This approach was used by Silvey [42] to assign sucrose octasulfate. Although such methods are not as direct as multi-dimensional methods for unambiguous assignment, they may help with assignments when access to more sophisticated instrumentation is limited.

Assignment of diastereotopic methylene protons in carbohydrates, such as those on C-6 of the hydroxymethyl side group of hexopyranose rings, can be achieved either by selective deuteration [32–34] or by Karplus analysis of  $^3J_{C4,H6}$  and  $^3J_{H5,H6}$  [43–46].

Combining selective excitation methods to generate subspectra of each sugar residue, such as 1D HOHAHA (see below), with comparison of structural reporter resonances can streamline the analysis [47]. But definitive assignments require a multi-dimensional approach.

## ***2.2. Two-dimensional methods for assignment: homonuclear and heteronuclear***

Information between nuclear spins can be transferred either coherently or incoherently (see Chapter 1). Scalar coupling transfers magnetization coherently, whereas dipolar interactions that lead to the nuclear Overhauser effect transfer magnetization incoherently. Correlation spectroscopy is the term applied to methods that exploit scalar coupling between like ( $^1H$ – $^1H$ ,  $^{13}C$ – $^{13}C$ ) or unlike ( $^1H$ – $^{13}C$ ,  $^1H$ – $^{31}P$ ) nuclei. If one resonance is assigned, then assignment of its coupling partner can be made by direct inspection of correlation spectra.

Any pulse sequence that exploits connectivities between  $^1H$  and  $^{13}C$  can be applied to carbohydrates. Useful methods for assignment within each sugar residue exploit scalar coupling among ring protons and scalar coupling between a ring carbon and its attached proton. Standard pulse sequences such as double quantum filtered COSY (DQCOSY; [48]), total correlation spectroscopy (TOCSY; [49–51]), and HMQC [52] will usually suffice to get most  $^1H$  and  $^{13}C$  assignments, but complete and unambiguous assignment may require additional, less routine methods.

In some cases, correlations between  $^1H$  and other heteronuclei may be exploited, either by natural occurrence or by insertion of a heteronuclear probe. For example,  $^1H$ – $^{31}P$  correlation can be used for assignment of deoxyribose or ribose ring protons in DNA [53] or RNA [54], while  $^{31}P$ – $^1H$  coupling constants provide information about the backbone

geometry of oligonucleotides [55–57]. Reddy and coworkers [58] used scalar coupling between  $^{31}\text{P}$  and  $^{13}\text{C}$  to locate the position of a phosphodiester linkage in a bacterial cell wall polysaccharide.  $^{19}\text{F}$  is a useful probe when it can be inserted synthetically into carbohydrates [59–62]. In general, the wealth of multi-dimensional methods utilizing correlations with  $^{15}\text{N}$  that are available for the analysis of proteins (see Chapters 2 and 5) are not applicable to carbohydrates for the obvious reason that nitrogen is not a common constituent.

### 2.2.1. *Homonuclear correlation spectroscopy*

COSY and its variants are useful for delineating coupled networks of protons within each sugar residue. DQCOSY is usually preferable to a regular COSY for oligosaccharides because of suppression of any residual water signal and improved resolution near the diagonal. COSY spectra may be desirable in addition to a TOCSY spectrum because they may have fewer crosspeaks and direct coupling partners are unambiguously identified. Triple-quantum-filtered COSY (TQF-COSY) experiments, which select for coupling networks involving three coupled protons, are useful for assigning protons at positions 5 and 6 of a pyranose residue [63]. Note that phase-sensitive versions [64,65] of COSY experiments are required to achieve the resolution necessary for carbohydrates.

The homonuclear Hartman–Hahn (HOHAHA) experiment of Bax and Davis [51], the most widely used version of total correlation spectroscopy (TOCSY), is often the most useful homonuclear correlation method because it reveals complete scalar coupling networks within each subunit of an oligosaccharide. Starting from a known resonance, usually an anomeric proton, crosspeaks belonging to the same subunit in an oligosaccharide can be identified. However, intensities of HOHAHA crosspeaks must be interpreted with care as they are complicated functions of all the  $J$  values involved [66–69]. Quantitative  $^3J_{\text{HH}}$  values can be extracted from HOHAHA experiments by iterative back-calculation of cross peak intensities based on the full Hamiltonian operating during the pulse sequence [68].

Since the introduction of the HOHAHA experiment, many variations to increase sensitivity [70,71], suppress cross-relaxation [72–78]; or improve resolution [79,80] have been suggested. For oligosaccharides with well-separated anomeric resonances, the one-dimensional version [79] with selective excitation offers better resolution than a full two-dimensional experiment. The addition of a  $z$ -filter can improve the accuracy of  $^3J_{\text{HH}}$

values derived directly from the splitting in a 1D HOHAHA spectrum [80,81].

One potential stumbling block in using HOHAHA for assignment arises when the transfer of magnetization is virtually halted by a small value of  $^3J_{\text{HH}}$ . Common constituents of oligosaccharides with small coupling constants are D-galactose ( $^3J_{\text{H4H5}} \sim 1$  Hz); D-mannose ( $^3J_{\text{H1H2}} \sim 1$ -2 Hz); and L-fucose ( $^3J_{\text{H4H5}} \sim 1$  Hz) [82]. For such cases, Inagaki and coworkers [83] suggested relayed-HOHAHA, a pulse sequence with MLEV17-based HOHAHA transfer, followed by the conventional pulse-interrupted free-precession method for relay between protons with small scalar couplings.

### *2.2.2. Heteronuclear correlation spectroscopy*

The greater dispersion of  $^{13}\text{C}$  chemical shifts makes  $^1\text{H}$ - $^{13}\text{C}$  chemical shift correlation spectroscopy an invaluable tool for assignment of oligosaccharides. The application to oligosaccharides of  $^1\text{H}$ -detected versions of one-bond and multiple-bond correlation pulse sequences has been described elsewhere [81]. Proton-detected heteronuclear multiple bond correlation (HMBC) [84], the most direct method for determining linkage patterns in oligosaccharides, correlates protons and carbons on either side of the glycosidic linkage. Although originally written to produce absolute value mode spectra, a newer version of HMBC utilizes "mixed mode" data processing to improve resolution and sensitivity [85]. For long-range correlations within a sugar ring, a 2D pulse sequence combining HMQC with HOHAHA [86] (sometimes called 2D HMQC-TOCSY or HMQC-HOHAHA) offers greater sensitivity than does HMBC.

### *2.3. Streamlining multi-dimensional experiments*

Marion and coworkers [87] demonstrated that high quality 2D spectra such as COSY, HOHAHA, and HMQC, could be recorded without phase cycling, thus reducing total acquisition times to a matter of minutes. The need for phase cycling is eliminated by using other means, including mode of data acquisition, to eliminate the most common spectral artifacts. This innovation makes 2D experiments suitable for rapid (5–10 minutes), routine assessment of small molecules. Application of these techniques renders 3D and 4D experiments on complex molecules such as proteins feasible.

Two, three- and four-dimensional experiments can be reduced by one dimension through the use of selective pulses [88]. This strategy is

especially appropriate for oligosaccharides which have relatively few protons, and clearly defined sub-systems of spins (each monosaccharide residue) which can usually be accessed via selective excitation of their respective anomeric protons. Reducing the dimensionality saves time and allows for better resolution in the acquisition domain (i.e., the equivalent of  $F_2$  in a 2D experiment). Numerous applications to oligosaccharides have appeared [12,89–93]. Selective excitation is usually achieved through the use of shaped pulses [94–96], but if the necessary spectrometer hardware is not available, it may be possible to use HOHAHA transfer from an anomeric proton to involve just one sugar ring of an oligosaccharide in a subsequent 2D experiment [97]. Kupe and Freeman [98] used multi-step HOHAHA transfers to measure selected  $^1\text{H}$   $T_1$  values in overlapping spectra. When using shaped pulses to obtain quantitative results, such as NOE intensities, it is important to keep the pulse length short relative to relaxation rates [99,100]. This is not a difficult criterion to meet for most oligosaccharides, especially if the selective excitation is being applied to a well-separated anomeric proton so that a relatively short shaped pulse can be used. (The selectivity of the pulse is qualitatively proportional to its length.)

The application of pulsed field gradients (PFG) dramatically shortens acquisition times for multi-dimensional experiments [101,102] and applies as well to oligosaccharides as to any other sample [103]. In fact, sucrose or glucose often appears as the test sample for new pulse sequences using gradients [101,104–106]. The principle is to select desired coherence pathways and destroy unwanted pathways by the application of field gradients instead of by phase cycling. The elimination of phase cycling results in a tremendous savings in time provided the sample is concentrated enough to yield sufficient signal-to-noise in a single scan. This condition may not always be met for  $^1\text{H}$ -detected heteronuclear correlation experiments on dilute solutions of hard-to-get oligosaccharides.

PFG spectroscopy can also be used for better solvent suppression than traditional presaturation because protons of  $\text{H}_2\text{O}$  usually only generate single-quantum coherences [101]. As an added benefit, the full dynamic range of the ADC can be used for resolving desired signals because the huge signal from water is eliminated before the receiver. The application of PFGs should aid observation of hydroxyl protons of oligosaccharides in aqueous solutions, especially those closer to the water resonance. Gradients can also be used to suppress the water signal by exploiting the difference in diffusion rates for solvent mole-

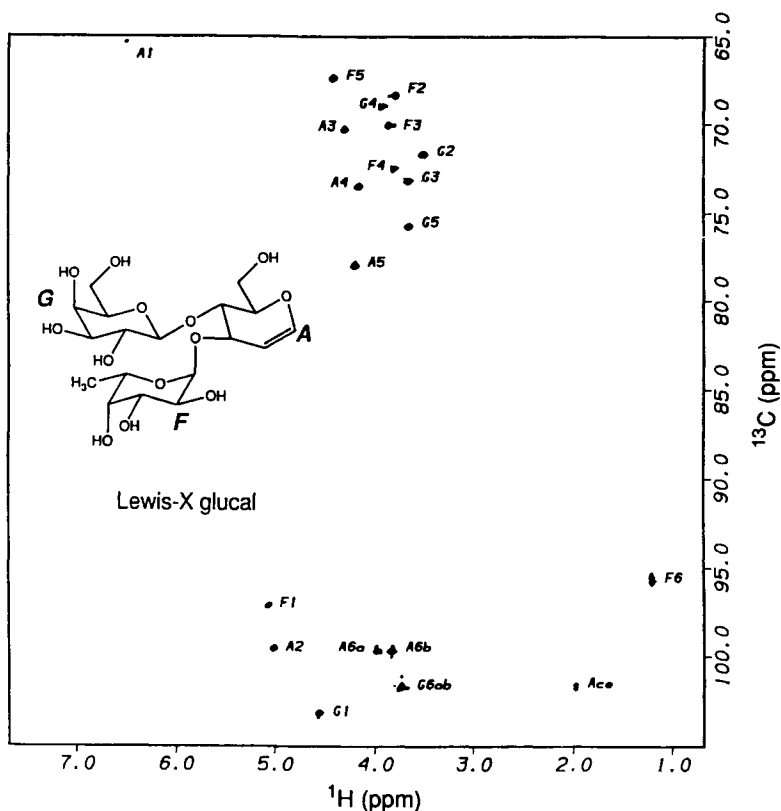


Fig. 2. Phase-sensitive HMQC spectrum at 500 MHz of Lewis<sup>x</sup> glucal, approximately 40 mM in D<sub>2</sub>O, four scans per  $t_1$  value, 512  $t_1$  values (256 complex values). The total measuring time was 36 minutes. The peak labelled "Ace" at 2.0 ppm ( $^1\text{H}$ ) is due to traces of acetamide present in the sample. From Ref. [111], by permission.

cules and sample molecules [107]. For  $^1\text{H}$ -detected heteronuclear correlation experiments, gradients offer much better suppression of  $^1\text{H}$  resonances bound to  $^{12}\text{C}$  or  $^{14}\text{N}$  [108–110]. To generate pure-phase HMQC spectra, Tolman and coworkers [111] used a gradient that alternated sign with each  $t_1$  value immediately after the final  $^{13}\text{C}$  90° pulse. Application to an oligosaccharide is shown in Fig. 2. The use of PFGs to improve  $^1\text{H}$ -detected heteronuclear correlation experiments is especially attractive when combined with isotopic labelling to improve S/N and allow detection of  $^{13}\text{C}$ – $^{13}\text{C}$  correlations [112].



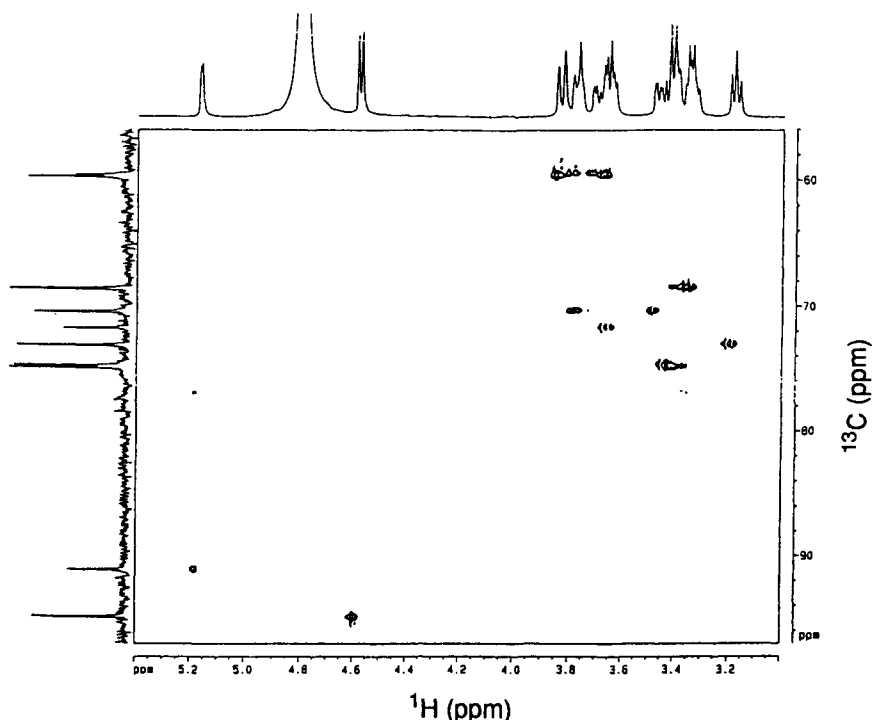


Fig. 3. PMG-HMQC spectrum at 500 MHz of an equilibrium mixture of  $\alpha$ - and  $\beta$ -D-glucose, 200 mM in  $D_2O$ , recorded with two scans per  $t_1$  value and 256  $t_1$  values. The total measuring time was approximately 18 minutes. From Ref. [115], with permission.

The practical drawback for implementation of PFGs is the expense of the required spectrometer accessories. As an alternative, Canet and others [104–106,113,114] demonstrated that  $B_1$  (radiofrequency) gradients could also be used for coherence selection. The hardware required is a single-turn coil added to a conventional NMR probe, and a second transmitter. An even more economical alternative, the PMG (poor man's gradient) method, was suggested by Nuzillard and coworkers [115] for eliminating phase cycling in the HMQC experiment. They inserted a spin-lock purging pulse [116] in the standard BIRD-pulse HMQC sequence [52] to improve suppression of signals from  $^1H$  nuclei not bound to  $^{13}C$ . This spin-lock pulse acts as a rotating-frame  $B_1$  gradient pulse because of  $B_1$  field inhomogeneities. An example of the effectiveness of this pulse sequence is shown in Fig. 3.



derived from crosspeak intensities obtained from conventional 2D NOESY spectra with intensities obtained from 3D and 2D NOESY-HOHAHA and HMQC-NOESY spectra. Calculated interproton distances, both intra- and inter-residue varied by approximately  $\pm 0.3$  Å depending on the pulse sequence used to generate the crosspeak. These results illustrate the difficulty in using three-dimensional methods to quantify NOEs [127]. However, 3D methods, or their 2D or 1D analogs, such as TOCSY-NOESY and TOCSY-ROESY [90,128] are useful for determining which of two or more overlapping spins are contributing to a particular NOE crosspeak [90,91].

### 3. CONFORMATION

After primary structure, the next question is conformation. Conformational analysis of oligosaccharides uses the same basic approach as for proteins or nucleic acids, but with fewer tools available. Besides the overlap in the proton spectrum, the dearth of nitrogen nuclei, and the difficulty of isotopic labelling, the most serious obstacle in using NMR for conformational analysis is the essentially linear structure of oligosaccharides. Lemieux and coworkers, among others, demonstrated that the conformation of an oligosaccharide is characterized mainly by the  $\phi$  and  $\psi$  angles that describe torsion about the glycosidic linkage [129]. (A third torsion angle,  $\omega$ , is needed to describe 1,6 linkages.) Typically, the only constraints on these angles that can be obtained are one or two interglycosidic NOEs and two long-range heteronuclear coupling constants. For the determination of the conformation of a protein, the accuracy of any particular NOE buildup rate or intensity is not that critical because hundreds of such values from neighboring proton pairs can be included, providing internal checks. For an oligosaccharide, no such cross-checks are available. Nevertheless, NOEs in conjunction with coupling constants can narrow the range of possible conformations.

#### 3.1. *Coupling constants*

Three-bond scalar coupling constants provide information about local conformation because the coupling depends on the dihedral angle with a dependence described originally by Karplus [130,131]. In general, for two vicinal protons on a CHCH fragment in a pyranoside which are oriented equatorial–equatorial,  $^3J_{HH}$  equals ~2-3 Hz; for axial–equa-

torial orientation  $^3J_{HH}$  equals ~2–6 Hz; and for axial–axial orientation,  $^3J_{HH}$  equals ~9–12 Hz [82,132]. Altona and Haasnoot [82] developed an equation for calculating accurate values for specific monosaccharides, based on the electronegativity and geometry of constituents. Values of  $^3J_{HH}$  for most pyranose rings of biological importance show that the rings are relatively fixed in the  $^4C_1$  conformation [133]. An important exception is iduronate found in heparin, heparan sulfate, and dermatan sulfate.  $^3J_{HH}$  values for idose and its derivatives are intermediate relative to the values expected for chair ( $^1C_4$  or  $^4C_1$ ) and skew-boat conformations. The values also depend on the neighboring residues in an oligosaccharide [134] and vary with solution conditions such as solvent and temperature [135]. These observations have been interpreted in terms of rapid conformational averaging between two or three ring conformations [134–137].

$^3J_{H5,H6R}$  and  $^3J_{H5,H6S}$  values can be used to estimate the relative populations of three possible rotamers about the C5–C6 bond in hydroxymethyl groups ([138,139] and references therein). However, this approach assumes three distinct rotamers about the C5–C6 bond, rather than allowing for a continuous distribution [140]. Recently, Poppe [139] used the maximum entropy method [141] to allow for the possibility of a continuous distribution in  $\beta$ -gentiobiose, and concluded that there were two major rotamer populations.

There are numerous methods available for measuring long-range interglycosidic  $^3J_{COCH}$  values ([142] and references therein). Recently, Poppe and coworkers were able to measure  $^3J_{COCH}$  values by a  $^{13}C$ - $\omega_1$ -half-filtered  $^1H$ - $^1H$  ROESY sequence [143]. What is not yet available is a definitive Karplus relation between these values, usually ~4–6 Hz, and  $\phi$  and  $\psi$  values, although some forms have been suggested [144,145].

As discussed below, enrichment with  $^{13}C$  enables measurement of one-bond and long-range  $^{13}C$ - $^{13}C$  coupling constants. By means of selective labelling with  $^{13}C$  by chemical and/or enzymatic methods, Serianni and coworkers are developing rules for interpreting a wide variety of homo- and heteronuclear  $^{13}C$  coupling constants in terms of ring and interglycosidic conformations in five- and six-membered rings [146–149]. For example, they have shown that  $^1J_{CC}$  values of HO–C–C–OH fragments in aldofuranoses and aldopyranoses are sensitive to the dihedral angle [146].  $^2J_{CCC}$  and  $^3J_{CCCC}$  values are sensitive to the orientation of substituents and dihedral angles in both furanose and pyranose rings [147]. Tvaroska and Taravel [150] have proposed a Karplus-type relationship for  $^1J_{CH}$  values in carbohydrates.

### 3.2. Strong coupling

Strong coupling, when the chemical shift difference between nuclei is comparable to their  $^3J_{HH}$  value, complicates analysis of  $^1H$  spectra. It can lead to large changes in chemical shift with temperature [151,152], which could be misinterpreted as evidence of large conformational changes. It also complicates quantification of NOE cross peaks [153, 154]. The possibility of strong coupling should always be considered when measuring  $^1H$  chemical shift temperature coefficients or  $^1H$ - $^1H$  NOEs for oligosaccharides with closely overlapped  $^1H$  resonances. It is possible to avoid some strong coupling effects among  $^1H$  resonances by transferring proton magnetization to attached  $^{13}C$  nuclei [155] and measuring  $^1H$   $T_1$ s indirectly. Going to higher field strengths will lessen the problem of strong coupling.

### 3.3. NOESY and ROESY

Typical interproton distances for proton pairs in three common pyranoses in the  $^4C_1$  conformation are listed in Table 1.

As can be inferred from Table 1, particular residues in an oligosaccharide will have characteristic patterns of strong and weak NOE intensities. For example,  $\beta$ -D-glucose is expected to exhibit a strong NOE between H2 and H4, whereas for  $\alpha$ -D-mannose or  $\beta$ -D-galactose, that NOE would be relatively weak or absent. As previously noted,  $^3J_{HH}$  values are consistent with relatively rigid  $^4C_1$  conformations for most pyranose rings. Therefore intraresidue NOEs or ROEs can be used to

TABLE 1

Interproton distances in Å obtained from macromodel 3.5 [156]

Monosaccharide proton pair	$\beta$ -D-glucose	$\alpha$ -D-mannose	$\beta$ -D-galactose
1,2	3.076	2.562	3.075
1,3	2.634	3.873	2.641
1,4	4.019	3.910	4.138
1,5	2.553	3.739	2.545
2,3	3.083	2.436	3.084
2,4	2.617	3.790	3.801
2,5	4.025	3.998	4.004
3,4	3.086	3.088	2.457
3,5	2.651	2.497	2.576
4,5	3.088	3.090	2.447

generate reference distances for the determination of interresidue distances. Also, intraresidue NOEs or ROEs can be used for assignment within a residue, assuming the conformation (e.g.,  ${}^4C_1$ ) is known from independent information (such as  ${}^3J_{HH}$  values). Inter-residue NOEs or ROEs can also be used to confirm the sequence of residues. Note that many oligosaccharides of interest, being of molecular weight  $\sim 1000$  Daltons, may exhibit such small NOEs that a ROESY (rotating frame NOE spectroscopy) experiment is usually preferable. (ROE crosspeaks are always positive, whereas the sign and intensity of NOE crosspeaks depend on the correlation time of the internuclear vector, and hence on the molecular weight of the molecule [157,158]). In fact, the introduction of rotating frame NOE measurement, the CAMELSPIN experiment, was illustrated by its application to a tetrasaccharide [158].

### ***3.4. Using relaxation parameters to characterize flexibility***

The standard approach to defining the solution conformation of an oligosaccharide is to use interglycosidic NOEs and  ${}^3J_{COCH}$  values to determine a single conformation (set of  $\phi$ ,  $\psi$  values). These values are then compared to those obtained from theoretical conformations calculated based on some force field. The three-dimensional potential energy surface, plotted as a function of  $\phi$  and  $\psi$  values, may indicate a narrow or broad region of minimal potential energy. A detailed discussion of computer modelling of carbohydrates is beyond the scope of this chapter but many recent examples and reviews are available [159–173].

A broad minimum and/or a broad range of  $\phi$ ,  $\psi$  values compatible with the NMR-derived constraints suggests the existence of multiple conformers. Whether oligosaccharides exist in solution as single conformers or equilibrating mixtures of several conformers is still being debated [174]. By using a model disaccharide,  $\alpha$ -D-Man(1 $\rightarrow$ 3)- $\alpha$ -D-Man-OMe, Rutherford and coworkers [88] demonstrated that both a single, rigid conformation and an average among conformations could fit the data (three NOEs and two long-range coupling constants) equally well. Therefore, to detect flexibility in oligosaccharides, it is necessary to add  ${}^1H$  and  ${}^{13}C$  relaxation parameters to the arsenal [174–177].

The basic principles underlying the dependence of relaxation rates on molecular motion are described in Chapter 5. The same principles apply in oligosaccharides, of course, with the caveat that small oligosaccharides may have overall correlation times that place them close to the  $T_1$  minimum, especially at higher fields.

Several groups have demonstrated that particular disaccharides or oligosaccharides must be undergoing local motions about their glycosidic linkages, superimposed on the overall molecular tumbling of the molecule [88,175,178,179]. In general, these studies demonstrate that more than one low-energy conformation is consistent with NOE or relaxation data. In several cases, relaxation parameters were shown to be field-dependent, and this dependence was interpreted as the result of incomplete averaging of local motions occurring on about the same timescale as the overall molecular tumbling [175,178,179].

This similarity in timescales for local and global motions is another consequence of the relatively small size of oligosaccharides, and makes it impossible to interpret measured order parameters [181,182] in terms of internal motions about the glycosidic linkage. It may be possible to separate internal and global motions by putting the oligosaccharide in a more viscous solvent, tethering it to a larger molecule or inserting it into a membrane (see examples below), studying a larger fragment (if it is from a polysaccharide), and/or lowering the temperature.

Tethering the oligosaccharide to a slow-moving object may also be a more physiologically relevant state for the oligosaccharide if it is normally part of a glycoconjugate and/or attached to the cell membrane. Rutherford and coworkers [88] measured  $^{13}\text{C}$  relaxation parameters for an N-linked oligosaccharide covalently linked to ribonuclease B. They observed significant motion about the glycosidic linkages which increased along the chain as distance from the protein increased.

Aubin and coworkers [183,184] prepared sialyl residues with strategic  $^{13}\text{C}$  labels, synthetically incorporated them into ganglioside GM3, and then measured dipolar interactions between  $^{13}\text{C}$ – $^{13}\text{C}$  and  $^1\text{H}$ – $^{13}\text{C}$  pairs when the ganglioside was inserted in oriented bilayer systems. Poppe and coworkers [185] were able to measure proton-detected  $^{13}\text{C}$   $T_1$ ,  $T_{1\rho}$ , and  $^1\text{H}/^{13}\text{C}$  and  $^1\text{H}/^1\text{H}$  NOE values for the carbohydrate headgroup of a ganglioside by anchoring it in a perdeuterated dodecylphosphocholine micelle.

#### 4. ANALYSIS OF GLYCOCONJUGATES

Some studies using solid state  $^2\text{H}$ –NMR to study oligosaccharides (as glycoconjugates) in bilayers have been published [186–190]. Reports of high resolution  $^1\text{H}$  NMR studies of intact glycopeptides and glycoproteins are relatively rare, for the obvious reason that the spectra of glycoconjugates will be even more complex than of the glycosyl or

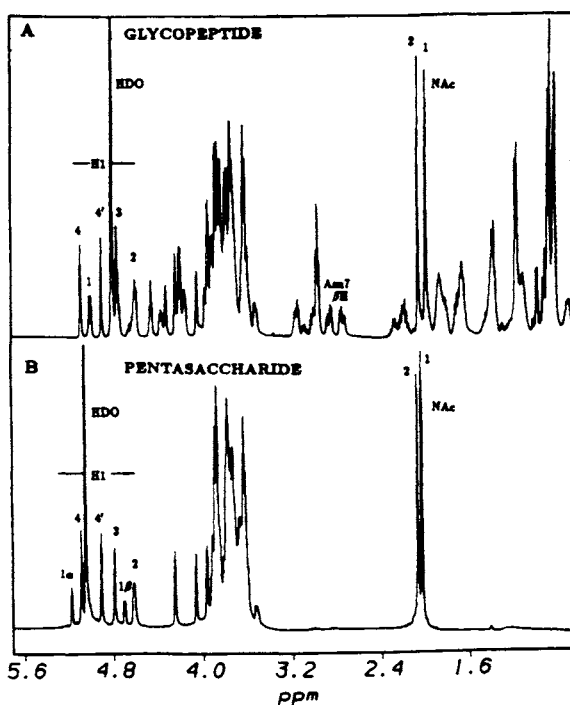
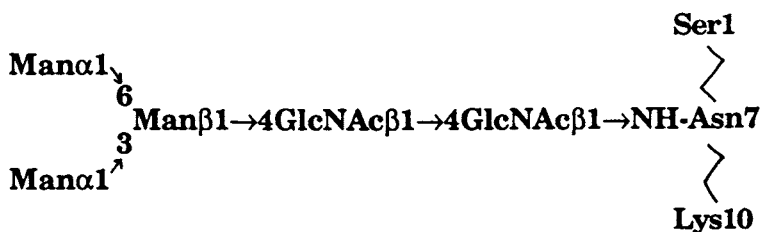


Fig. 4. (A) One-dimensional  $^1\text{H}$  NMR spectrum at 500 MHz of the glycopeptide in  $\text{D}_2\text{O}$  (50 mM phosphate, pD 3.0) at  $25^\circ\text{C}$ . The anomeric protons and GlcNAc methyl groups are indicated. (B) One-dimensional  $^1\text{H}$  NMR spectrum at 500 MHz of the pentasaccharide in  $\text{D}_2\text{O}$  at  $5^\circ\text{C}$ . The decapeptide is shown schematically. Adapted from Ref. [191], with permission.

aglycon moieties individually. For example, Figure 4 compares one-dimensional  $^1\text{H}$  spectra for a pentasaccharide with and without an N-linked decapeptide from hen ovomucoid [191].



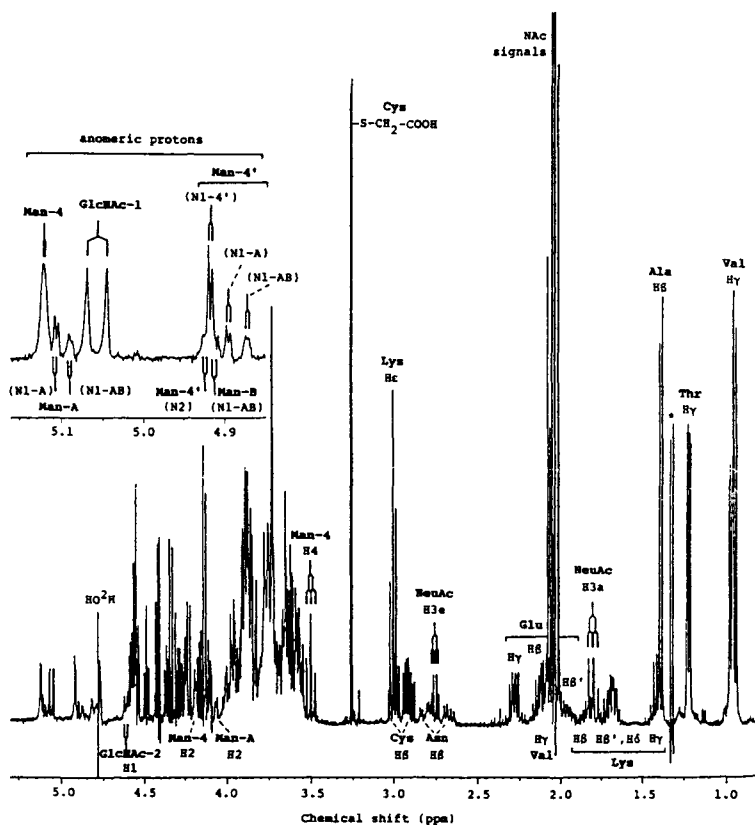


Fig. 5. 400 MHz  $^1\text{H}$  NMR spectrum of  $\alpha\text{GP-1}$ , a tryptic fragment of hGC, at 298 K. The sample was a mixture of four glycopeptides with identical peptide sequences but different oligosaccharides. Part of the anomeric region is expanded. The asterisk at  $\sim 1.33$  ppm is assigned to an impurity. From Ref. [192], with permission.

Two other examples of analyses of glycoconjugates by high resolution NMR are shown in Figs. 5 and 6. Weisshaar and coworkers [192] used one- and two-dimensional  $^1\text{H}$  NMR methods to analyze glycopeptides derived from human chorionic gonadotrophin (hCG). The types of oligosaccharides, their attachment points, and relative distribution along the hCG chain were assessed using tryptic and chymotryptic fragments varying in length from seven to sixteen amino acid residues. Each fragment carried one N-linked oligosaccharide, and considerable heterogeneity was found for these oligosaccharides. A typical one-dimensional  $^1\text{H}$

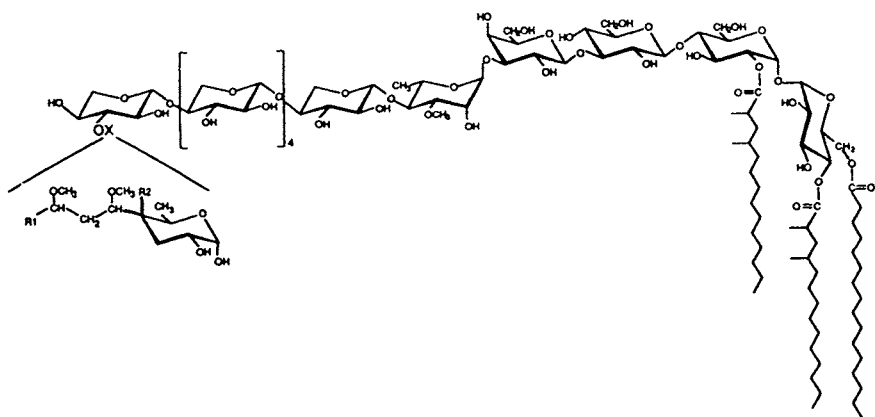


Fig. 6a. *Mycobacterium gastri* LOS-III structure ( $R_1$  and  $R_2$  are undetermined).

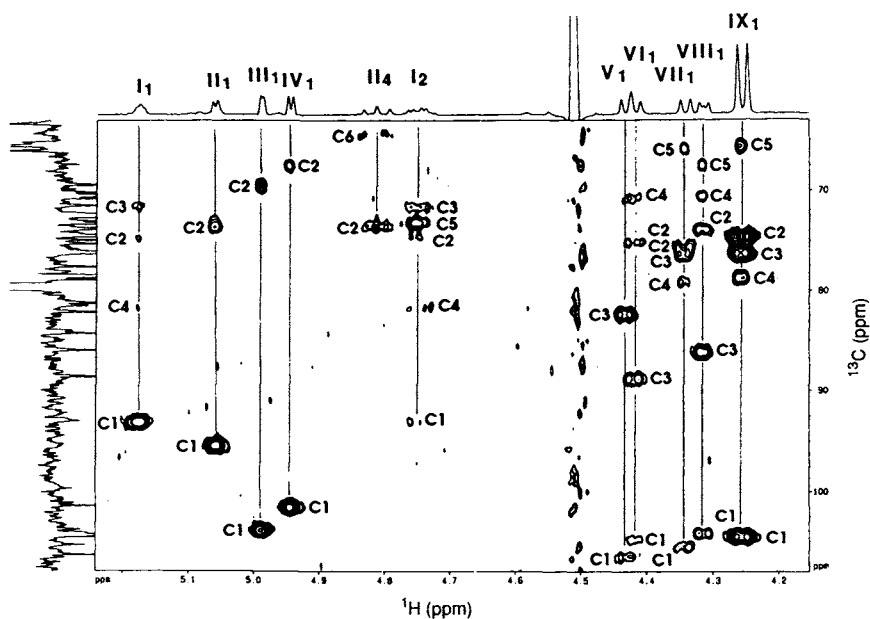


Fig. 6b. Partial 2D HMQC-HOHAHA spectrum of native LOS-III (lipo-oligosaccharide III) from *Mycobacterium gastri*, 10 mg in  $CDCl_3/CD_3OD$  (1:1, v/v) at 500 MHz. The data are presented in absolute value mode. Labels along the  $^1H$  chemical shift axis correspond to the nine distinct sugar residues in the structure, above. Adapted from Ref. [196], with permission.

NMR spectrum at 400 MHz of one of these fragments, assigned primarily by the structural reporter approach, is shown in Fig. 5. The sample consisted of mainly four glycopeptides with identical peptide moieties (twelve amino acid residues) and four different oligosaccharides.

In a few cases, researchers have been able to extract high resolution NMR data from oligosaccharides covalently linked to proteins. Rutherford and coworkers [88] measured  $^{13}\text{C}$  T<sub>1</sub>s via  $^1\text{H}$ -detection for an oligosaccharide N-linked to ribonuclease B. Withka and coworkers [193] tentatively identified some NOE crosspeaks between the adhesion domain of human T-lymphocyte glycoprotein CD2 and its attached oligosaccharide. Complicated spectra could be simplified either by attaching labelled sugars to unlabelled aglycons via semi-synthetic methods [184] or by pulse sequences that are "tailored" for specific regions of chemical shifts [194,195].

NMR is extremely valuable for identification of primary structure in lipopoly- and lipooligosaccharides, important components of bacterial cell walls which have been implicated in infection and resistance to infection. These compounds are analyzed by  $^1\text{H}$  NMR in mixtures of organic solvents such as methanol and chloroform, because of the limited solubility of a molecule that is partly polar (carbohydrate) and partly nonpolar (lipid). This poor solubility also may restrict analysis to  $^1\text{H}$  NMR because of signal-to-noise limitations, although it is possible to use reverse-detected methods to obtain  $^{13}\text{C}$  parameters. Gilleron and coworkers used a combination of COSY, HOHAHA, HMQC, HMBC, and HMQC-HOHAHA spectroscopy to identify a novel C4-branched lipo-oligosaccharide from *Mycobacterium gastri*, a model for mycobacterial pathogens implicated in human diseases such as tuberculosis [196].

Nature is not required to restrict itself to combining only two types of biomolecules, so perhaps it is inevitable that there would be a need to analyze glycopeptidolipid (GPL) samples. Rivière and coworkers [197] used COSY, relayed COSY, HMQC, HMBC, HMQC-HOHAHA, and J-modulated one-dimensional  $^{13}\text{C}$  spectroscopy to assign  $^1\text{H}$  and  $^{13}\text{C}$  resonances for a GPL isolated from *Mycobacterium xenopi*.  $^3J_{\text{HH}}$  values were extracted either from one-dimensional  $^1\text{H}$  spectra or from phase-sensitive DQCOSY spectra. This GPL was found to contain two lipid chains, three glycosyl units, and four amino acid residues.

Since cell surface glycoproteins are critical for cell recognition events, it is important to explore the behavior of their oligosaccharide components in cell membranes. It may be that a particular pattern of clustered oligosaccharides are necessary for recognition. Methods must be

developed for studying the pattern in the carpet of carbohydrates on the cell surface, not just the behavior of an isolated strand [1,198–200]. This is a difficult challenge. NMR offers some approaches using model systems. Solid state NMR methods used to study membrane proteins could be applied to glycoproteins provided there is some way to label the glycosidic moieties with  $^2\text{H}$  [201]. It is possible to obtain high resolution  $^1\text{H}$  spectra using model membrane systems. Some examples were cited in the section on measuring relaxation parameters of oligosaccharides [183–185].

## 5. INTERACTIONS OF OLIGOSACCHARIDES

Interactions between an oligosaccharide and a large receptor, such as an antibody, potentially can be studied using the transferred NOE experiment (TRNOE) [202,203]. In this experiment, the difference in correlation time between a free and bound small ligand is exploited. When bound to a large receptor, the small ligand rapidly develops intramolecular NOEs because of the long correlation time of the bound complex. If the exchange rate between bound and free is sufficiently high,  $^1\text{H}$  linewidths in the acquired NOESY spectra will be relatively narrow, reflecting the correlation time of the free ligand. By measuring the NOEs in the absence and presence of the large receptor, a comparison of the free and bound ligand conformations can be made. Requirements and limitations of the method have been reviewed by Clore and Gronenborn [199,200] and Campbell and Sykes [204]. This approach has been applied successfully to several carbohydrate ligands [205–208]. As yet, there are too few examples to draw any general conclusions about how binding occurs and the role of flexibility of oligosaccharides in their recognition by proteins. Bevilacqua and coworkers [205,206] found that Ricin B bound the major conformer of  $\alpha\text{-D-Gal}(1\rightarrow6)\beta\text{-D-GlcOMe}$  in preference to the minor conformer. In contrast, the free conformation of a digalactose hapten was significantly different than when it was bound to an Fab fragment [207]. This suggests that the antibody either deforms the ligand upon binding or that it selects a minor conformer out of a distribution of conformers. For a third system, Weimar and Peters concluded that multiple conformations of a disaccharide could bind to an agglutinin [208]. Carver has discussed the thermodynamic implications of oligosaccharide flexibility and binding [209].

If a particular carbohydrate-receptor system is unsuitable for TRNOE analysis, it is still possible to study it by combining chemical modification of the carbohydrate ligand to determine which functional groups are necessary for binding, and NMR to determine the solution conformations of the variously modified ligands. For example, Bock and coworkers [210] combined  $^3J_{\text{HH}}$  and ROE values with molecular modeling to explain why GlcNAc-transferase V reacts only with one of two possible low-energy conformers of  $\beta$ -D-GlcNAc(1 $\rightarrow$ 2)- $\alpha$ -D-Man(1 $\rightarrow$ 6)- $\beta$ -D-GlcOR [211]. Another approach for exploring carbohydrate-protein interactions is to use fluorinated deoxy sugars to probe the role of hydroxyl groups in binding to proteins. Luck [61] has used this strategy to determine exchange rates, anomeric preferences, and the effect of temperature for the binding of monosaccharides to the D-glucose and D-galactose receptor protein.

## 6. FUTURE DIRECTIONS

Multidimensional methods with selective excitation make it feasible to assign  $^1\text{H}$  and  $^{13}\text{C}$  resonances of almost any oligosaccharide. Yet assignment in itself is only the first step in an NMR analysis. Assignment leads to identification of the sequence of residues in an oligosaccharide but, as previously discussed, NMR may not be the most efficient method for primary sequencing. However, NMR is a powerful method for determining the solution conformation of oligosaccharides under physiological conditions, and for studying their interactions with other molecules. NMR may be the only method available to study these aspects, if the system of interest will not crystallize.

If insertion of  $^{13}\text{C}$  or  $^2\text{H}$  labels into oligosaccharides could be done more readily, it would open a wide range of experiments [212]. Carbon-13 labels can greatly facilitate more accurate measurement of  $^1\text{H}$ - $^{13}\text{C}$  and  $^{13}\text{C}$ - $^{13}\text{C}$  coupling constants in oligosaccharides [9,212-215].  $^{13}\text{C}$  substitution can also allow the use of isotope editing or selective excitation to simplify spectra [216] and to study the conformation of a  $^{13}\text{C}$ -labelled ligand bound to an unlabelled host [217]. Given the relative ease of labelling proteins using expression systems, it may be simpler to study a fully-deuteriated protein with a protonated oligosaccharide ligand, as has been done for cyclosporin A and cyclophilin [218]. Selective deuteration can also be used with spectral editing to simplify spectra [219].

In principle, it should be possible to insert labels via *de novo* synthesis, synthesis starting with metabolically-labelled subunits, or some expression system, as has been done with great success for proteins. *In vitro* synthesis of carbohydrates using selective glycosyltransferases has the potential to allow incorporation of isotopically-labelled precursors under more controlled conditions than cell culture allows.

Historically, the attention of many biochemists was focussed on proteins and nucleic acids, while carbohydrates were considered as mainly structural or protective components. Carbohydrates are more difficult to isolate, purify, synthesize, and characterize, which may have contributed to their relative unpopularity. Despite these obstacles, much valuable progress was made by pioneering carbohydrate chemists before the advent of sophisticated multi-dimensional methods. It is hoped that this review has demonstrated that the increasing recognition of the important and diverse functions of carbohydrates, in combination with the technical advances described herein, is leading to even more rapid progress in our understanding of this class of biomolecules.

### **Acknowledgements**

I am grateful to J.H. Prestegard, John Ralph, A.S. Serianni, and Jun Shen for their helpful comments and advice. I also thank Sherry Naffz and Sharon Horstmann, who were instrumental in collecting references during the preparation of this manuscript.

### **REFERENCES**

1. A. Varki, *Glycobiology* 3 (1993) 97.
2. J.C. Lefebvre, V. Giordanengo, A. Doglio, L. Cagnon, J.P. Breittmayer, J.F. Peyron and J. Lesimple, *Virology*, 199 (1994) 265.
3. N.K.T. Back, L. Smit, J.J. Dejong, W. Keulen, M. Schutten, J. Goudsmit and M. Tersmette, *Virology*, 199 (1994) 431.
4. M. Haidar, J.C. Gluckman and L. Gattegno, *Glycobiology*, 2 (1992), 429.
5. E. Mbemba, J.C. Gluckman and L. Gattegno, *Glycobiology*, 4 (1994) 13.
6. D.A. Cumming, *Glycobiology*, 1 (1991) 115.
7. A. Dell and C. Jones, *Carbohydr. Res.*, 221 (1991) R7.
8. R.A. Dwek, C.J. Edge, D.J. Harvey, M.R. Wormald, and R.B. Prekh, *Annu. Rev. Biochem.*, 62 (1993) 65.
9. A.S. Serianni, in: *Glycoconjugates: Composition, Structure and Function* (H.J. Allen and E.C. Kisailus, eds.). Marcel-Dekker, 1992, p. 71.

10. J. Dabrowski, in: *Methods in Stereochemical Analysis: Two-Dimensional NMR Spectroscopy* (W.R. Croasmun and R.M.K. Carlson, eds.). VCH Publishers, Inc., 1987, p. 349.
11. H. van Halbeek, in: *Frontiers of NMR in Molecular Biology*. Alan R. Liss, 1990, p. 195.
12. S.W. Homans, *Prog. NMR Spectros.*, 22 (1990) 55.
13. B.R. Leeftang, Ph.D. thesis, Utrecht University, 1991.
14. H. van Halbeek and L. Poppe, *Magn. Reson. Chem.*, 30 (1992) S74.
15. P.K. Agrawal, *Phytochemistry*, 31 (1992) 3307.
16. E.C. Kisailus and H.J. Allen, in: *Glycoconjugates: Composition, Structure and Function* (H.J. Allen and E.C. Kisailus, eds.). Marcel-Dekker, 1992, p. 13.
17. R.U. Lemieux and J.D. Stevens, *Can. J. Chem.*, 44 (1966) 249.
18. D.R. Bundle and R.U. Lemieux, *Meth. Carbohydr. Chem.*, 7 (1976) 79.
19. J.F.G. Vliegthart, H. van Halbeek and L. Dorland, *Pure Appl. Chem.*, 53 (1981) 45.
20. J.F.G. Vliegthart, L. Dorland and H. van Halbeek, *Adv. Carbohydr. Chem. Biochem.*, 41 (1983) 209.
21. R. Barker, H.A. Nunez, P. Rosevear and A.S. Serianni, *Meth. Enzymol.* 83 (1982) 58.
22. K. Bock and C. Pedersen, *Adv. Carbohydr. Chem. Biochem.*, 41 (1983) 27.
23. B. Adams and L.E. Lerner, *Magn. Reson. Chem.*, 32 (1994) 225.
24. L. Poppe and H. van Halbeek, *J. Am. Chem. Soc.*, 113 (1991) 363.
25. B. Adams and L. Lerner, *J. Magn. Reson.*, 96 (1992) 604.
26. G. Okide, D.S. Weber and W.J. Goux, *J. Magn. Reson.*, 96 (1992) 526.
27. J.A. van Kuik, K. Hård and J.F.G. Vliegthart, *Carbohydr. Res.*, 235 (1992) 53.
28. D.S. Weber and W.J. Goux, *Carbohydr. Res.*, 233 (1992) 65.
29. K. Hermansson, P.-E. Jansson, L. Kenne, G. Widmalm and F. Lindh, *Carbohydr. Res.*, 235 (1992) 69.
30. G.V.T. Swapna and R. Ramachandran, *J. Magn. Reson. Series B*, 102 (1993) 204.
31. S.J.E. Taylor, R.A. Rastall, J.T. Sykes, S.C. Winter and C. Bucke, *Comp. Appl. Biosci.*, 9 (1993) 723.
32. D. Gagnaire, D. Horton and F.R. Taravel, *Carbohydr. Res.*, 27 (1973) 363.
33. H. Ohrui, H. Horiki, H. Kishi and H. Meguro, *Agric. Biol. Chem.*, 46 (1983) 1101.
34. H. Ohrui, Y. Nishida, H. Itoh and H. Meguro, *J. Org. Chem.*, 56 (1991) 1726.
35. K. Akasaka, T. Imoto, S. Shibata and H. Hatano, *J. Magn. Reson.*, 18 (1975) 328.
36. P. Dais and A.S. Perlin, *Adv. Carbohydr. Chem. Biochem.*, 45 (1987) 125.
37. H. Hori, H. Ohrui, H. Meguro, Y. Nishida and J. Uzawa, *Tetrahedron Lett.*,

- 29 (1988) 4457.
38. P.C. Kline, A.S. Serianni, S.G. Huang, M. Hayes and R. Barker, *Can. J. Chem.*, 68 (1990) 2171.
39. T. Katsura, K. Ueno and K. Furusawa, *Magn. Reson. Chem.*, 31 (1993) 1039.
40. L. Szilágyi and P. Forgó, *Carbohydr. Res.*, 247 (1993) 129.
41. P.E. Pfeffer, K.M. Valentine and F.W. Parrish, *J. Am. Chem. Soc.*, 101 (1979) 1266.
42. G.L. Silvey, *J. Pharm. Sci.*, 81 (1992) 471.
43. J.-R. Brisson and J.P. Carver, *Biochemistry*, 22 (1983) 3680.
44. P.C. Kline and A.S. Serianni, *Magn. Reson. Chem.*, 26 (1988) 120.
45. M.L. Hayes, A.S. Serianni and R. Barker, *Carbohydr. Res.*, 100 (1982) 87.
46. L. Poppe, W.S. York and H. van Halbeek, *J. Biomol. NMR*, 3 (1993) 81.
47. D. Machytka, R.A. Klein and H. Egge, *Carbohydr. Res.*, 154 (1994) 289.
48. U. Piantini, O.W. Sørensen and R.R. Ernst, *J. Am. Chem. Soc.*, 104 (1982) 6800.
49. L. Braunschweiler and R.R. Ernst, *J. Magn. Reson.*, 53 (1983) 521.
50. L. Braunschweiler and R.R. Ernst, *J. Magn. Reson.*, 53 (1985) 521.
51. A. Bax and D.G. Davis, *J. Magn. Reson.*, 65 (1985) 355.
52. A. Bax and S. Subramanian, *J. Magn. Reson.*, 67 (1986) 565.
53. V. Sklenář, A. Bax and G. Zon, *J. Am. Chem. Soc.*, 109 (1987) 2221.
54. G.W. Kellogg, A.A. Szewczak and P.B. Moore, *J. Am. Chem. Soc.*, 114 (1992) 2727.
55. V. Sklenář and A. Bax, *J. Am. Chem. Soc.*, 109 (1987) 7525.
56. A. Bax and L. Lerner, *J. Magn. Reson.*, 79 (1988) 429-438.
57. K.V.R. Chary, V.K. Rastogi, G. Govil and H.T. Miles, *J. Chem. Soc. Chem. Comm.*, 3 (1994) 241.
58. G.P. Reddy, C.-C. Chang and C.A. Bush, *Anal. Chem.*, 65 (1993) 913.
59. T.C. Wong, V. Rutar, J.-S. Wang, M. Feather and P. Ková, *J. Org. Chem.*, 49 (1984) 4358.
60. M.D. Percival and S.G. Withers, *Biochemistry*, 31 (1992) 505.
61. R.J. Abraham, E.J. Chambers and W.A. Thomas, *Magn. Reson. Chem.*, 32 (1994) 248.
62. L. Luck, in: *Techniques in Protein Chemistry VI* (J.W. Crabb, ed.). Protein Society, 1994.
63. S.W. Homans, R.A. Dwek, J. Boyd, M. Mahmoudian, W.G. Richards and T.W. Rademacher, *Biochemistry*, 25 (1986) 6342.
64. D. Marion and K. Wüthrich, *Biochem. Biophys. Res. Commun.*, 113 (1983) 967.
65. D.J. States, R.A. Haberkorn and D.J. Ruben, *J. Magn. Reson.*, 48 (1982) 286.
66. R. Bazzo and J. Boyd, *J. Magn. Reson.*, 75 (1987) 452.
67. C. Griesinger and R.R. Ernst, *Chem. Phys. Lett.*, 152 (1988) 239.



68. J.P.M. van Duynhoven, J. Goudriaan, C.W. Hilbers and S.S. Wijmenga, *J. Am. Chem. Soc.*, 114 (1992) 10055.
69. S.J. Glaser, *J. Magn. Reson. Series A*, 104 (1993) 283.
70. H.R. Bircher, C. Müller and P. Bigler, *J. Magn. Reson.*, 89 (1990) 146.
71. H. Bircher, C. Müller and P. Bigler, *Magn. Reson. Chem.*, 29 (1991) 726.
72. C. Griesinger, K. Wüthrich, G. Otting and R.R. Ernst, *J. Am. Chem. Soc.*, 110 (1988) 7870.
73. U. Kerssebaum, R. Markert, J. Quant, W. Bermel, S.J. Glaser and C. Griesinger, *J. Magn. Reson.*, 99 (1992) 184.
74. J. Cavanagh and M. Rance, *J. Magn. Reson.*, 88 (1990) 72.
75. J. Briand and R.R. Ernst, *Chem. Phys. Lett.*, 185 (1991) 276.
76. J. Cavanagh and M. Rance, *J. Magn. Reson.*, 96 (1992) 670.
77. J. Cavanagh and M. Rance, *J. Magn. Reson. Series A*, 105 (1993) 328 (correction).
78. M. Kadkhodaei, T.-L. Hwang, J. Tang and A.J. Shaka, *J. Magn. Reson. Series A*, 105 (1993) 104.
79. D.G. Davis and A. Bax, *J. Am. Chem. Soc.*, 107 (1985) 7197.
80. S. Subramanian and A. Bax, *J. Magn. Reson.*, 71 (1987) 325.
81. L. Lerner and A. Bax, *Carbohydr. Res.*, 166 (1987) 35.
82. C. Altona and C.A.G. Haasnoot, *Org. Magn. Reson.*, 13 (1980) 417.
83. F. Inagaki, D. Kohda, I. Shimada, A. Suzuki and A. Bax, *J. Magn. Reson.*, 81 (1989) 186.
84. A. Bax and M.F. Summers, *J. Am. Chem. Soc.*, 180 (1986) 2093.
85. A. Bax and D. Marion, *J. Magn. Reson.*, 78 (1988) 186.
86. L. Lerner and A. Bax, *J. Magn. Reson.*, 69 (1986) 375.
87. D. Marion, M. Ikura, R. Tschudin and A. Bax, *J. Magn. Reson.*, 85 (1989) 393.
88. H. Kessler, H. Oshkinat, C. Griesinger and W. Bermel, *J. Magn. Reson.*, 70 (1986) 106.
89. T.J. Rutherford, J. Partridge, C.T. Weller and S.W. Homans, *Biochemistry*, 32 (1993) 12715.
90. L. Poppe and H. van Halbeek, *J. Magn. Reson.*, 96 (1992) 185.
91. S. Holmbeck, P. Hajduk and L. Lerner, *J. Magn. Reson. Series B*, 102 (1993) 107.
92. H. Schröder and E. Haslinger, *Angew. Chem. Int. Ed. Engl.*, 32 (1993) 1349.
93. D. Uhrin, J.-R. Brisson and D.R. Bundle, *J. Biomol. NMR*, 3 (1993) 367.
94. L. Emsley and G. Bodenhausen, *Chem. Phys. Lett.*, 165 (1990) 469.
95. H. Geen and R. Freeman, *J. Magn. Reson.*, 93 (1991) 93.
96. L. Emsley and G. Bodenhausen, *J. Magn. Reson.*, 97 (1992) 135.
97. J.M. Nuzillard and G. Massiot, *J. Magn. Reson.*, 91 (1991) 380.
98. E. Kupce and R. Freeman, *J. Magn. Reson. Series A*, 101 (1993) 225.
99. P.J. Hajduk, D.A. Horita and L.E. Lerner, *J. Magn. Reson., Series A*, 103 (1993) 40.

100. D.A. Horita, P.J. Hajduk and L.E. Lerner, *J. Magn. Reson. Series A*, 103 (1993) 53.
101. R.E. Hurd, *J. Magn. Reson.*, 87 (1990) 422.
102. T.J. Norwood, *Chem. Soc. Rev.*, 23 (1994) 59.
103. A. Carpenter, L.D. Colebrook, L.D. Hall and G.K. Pierens, *Carbohydr. Res.*, 241 (1993) 267.
104. D.G. Cory, F.H. Laukien and W.E. Maas, *J. Magn. Reson. Series A*, 105 (1993) 223.
105. D. Canet, J. Brondeau, E. Mischler and F. Humbert, *J. Magn. Reson. Series A*, 105 (1993) 239.
106. W.E. Maas and D.G. Cory, *J. Magn. Reson. Series A*, 106 (1994) 256.
107. P. van Zijl and C.J. Moonen, *J. Magn. Reson.*, 87 (1991) 18.
108. B.K. John, D. Plant and R.E. Hurd, *J. Magn. Reson. Series A*, 101 (1992) 113.
109. G. Wider and K. Wüthrich, *J. Magn. Reson. Series B*, 102 (1993) 239.
110. Y.-C. Li and G.T. Montelione, *J. Magn. Reson. Series B*, 101 (1993) 315.
111. J.R. Tolman, J. Chung and J.H. Prestegard, *J. Magn. Reson.*, 98 (1992) 462.
112. J. Chung, J.R. Tolman, K.P. Howard and J.H. Prestegard, *J. Magn. Reson. Series B*, 102 (1993) 137.
113. J. Brondeau, D. Boudot, P. Mutzenhardt and D. Canet, *J. Magn. Reson.*, 100 (1992) 611.
114. W.E. Maas, F. Laukien and D.G. Cory, *J. Magn. Reson. Series A*, 103 (1993) 115.
115. J.-M. Nuzillard, G. Gasmi and J.-M. Bernassau, *J. Magn. Reson. Series A*, 104 (1993) 83.
116. G. Otting and K. Wüthrich, *J. Magn. Reson.*, 76 (1988) 569.
117. H.D. Plant, T.H. Mareci, M.D. Cockman and W.S. Brey, abstract A23, 27th Annual Experimental NMR Conference, Baltimore, MD, 1986.
118. C. Griesinger, O.W. Sørensen and R.R. Ernst, *J. Magn. Reson.*, 73 (1987) 574.
119. G.M. Clore and A.M. Gronenborn, *Science*, 252 (1991) 1390.
120. S.S. Wijmenga, K. Hallenga and C.W. Hilbers, *J. Magn. Reson.*, 84 (1989) 634.
121. S. Fesik and E.R.P. Zuiderweg, *J. Magn. Reson.*, 78 (1988) 588.
122. P. de Waard, B.R. Leeftang, J.F.G. Vliegthart, R. Boelens, G.W. Vuister and R. Kaptein, *J. Biomol. NMR*, 2 (1992) 211.
123. T.J. Rutherford and S.W. Homans, *Glycobiology*, 2 (1992) 293.
124. H. Schröder and E. Haslinger, *Liebigs Ann. Chem.*, 7 (1993) 751.
125. H. Schröder and E. Haslinger, *Liebigs Ann. Chem.*, 9 (1993) 969.
126. G.W. Vuister, P. de Waard, R. Boelens, J.F.G. Vliegthart and R. Kaptein, *J. Am. Chem. Soc.*, 111 (1989) 772.
127. R.R. Ernst, *Angew. Chem. Int. Ed. Engl.*, 31 (1992) 805.

128. D. Boudot, C. Toumestand, F. Toma and D. Canet, *J. Magn. Reson.*, 90 (1990) 221.
129. R.U. Lemieux and S. Koto, *Tetrahedron*, 30 (1974) 1933.
130. M. Karplus, *J. Chem. Phys.*, 30 (1959) 11.
131. M.J. Minch, *Concepts Magn. Reson.*, 6 (1994) 41.
132. B. Coxon, *Methods Carbohydr. Chem.*, 6 (1972) 513.
133. F.H. Cano and C. Foces-Foces, *J. Org. Chem.*, 52 (1987) 3361.
134. D.R. Ferro, A. Provasoli, M. Ragazzi, B. Casu, G. Torri, V. Bossennec, B. Perly and P. Sinay, *Carbohydr. Res.*, 195 (1990) 157.
135. P. Hajduk, Ph.D. thesis, University of Wisconsin, 1993.
136. A. Perlin, B. Casu, J. Tse and J.R. Sanderson, *Carbohydr. Res.*, 21 (1972) 123.
137. M.J. Forster and B. Mulloy, *Biopolymers*, 33 (1993) 575.
138. R.H. Marchessault and S. Pérez, *Biopolymers*, 18 (1979) 2369.
139. L. Poppe, *J. Am. Chem. Soc.*, 115 (1993) 8421.
140. D.A. Cumming and J.P. Carver, *Biochemistry*, 26 (1987) 6676.
141. E.T. Jaynes, in: *The Maximum Entropy Formalism* (R.D. Levine and M. Tribus, eds.). MIT Press, Cambridge, MA, 1979, p. 115.
142. B. Adams and L. Lerner, *J. Magn. Reson. Series A*, 103 (1993) 97.
143. L. Poppe, W.S. York and H. van Halbeek, *J. Biomol. NMR*, 3 (1993) 81.
144. A.S. Perlin and G.K. Hamer, in: *Carbon-13 NMR in Polymer Science* (W.M. Pasika, ed.). ACS Symposium Series, American Chemical Society, Washington, DC, 1979, p. 123.
145. I. Tvaroska, M. Hricovini and E. Petrakova, *Carbohydr. Res.*, 189 (1989) 359.
146. M.J. King-Morris and A.S. Serianni, *J. Am. Chem. Soc.*, 109 (1987) 3501.
147. J. Wu, P.B. Bondo, T. Vuorinen and A.S. Serianni, *J. Am. Chem. Soc.*, 114 (1992) 3499.
148. T. Bandyopadhyay, J. Wu and A.S. Serianni, *J. Org. Chem.*, 58 (1993) 5513.
149. J.M. Duker and A.S. Serianni, *Carbohydr. Res.*, 249 (1993) 281.
150. I. Tvaroska and F.R. Taravel, *Carbohydr. Res.*, 221 (1991) 83.
151. J.-R. Brisson and J.P. Carver, *J. Biol. Chem.*, 257 (1982) 11207.
152. W. Siciska, B. Adams and L. Lerner, *Carbohydr. Res.*, 242 (1993) 29.
153. J. Keeler, D. Neuhaus and M. Williamson, *J. Magn. Reson.*, 73 (1987) 45.
154. S.M.A. Holmbeck, P.A. Petillo and L.E. Lerner, *Biochemistry*, 33 (1994) 14246.
155. H.H. Lee, L.N. Congson, D.M. Whitfield, L.R. Radics and J.J. Krepinsky, *Can. J. Chem.*, 70 (1992) 2607.
156. F. Mohamadi, N.G.J. Richards, W.C. Guida, R. Liskamp, M. Lipton, C. Caulfield, G. Chang, T. Hendrickson and W.C. Still, *J. Comp. Chem.*, 11 (1990) 440.
157. S. Macura and R.R. Ernst, *Mol. Phys.*, 41 (1980) 95.
158. A.A. Bothner-By, R.L. Stephens, J.T. Lee, C.D. Warren and R.W. Jeanloz,

- J. Am. Chem. Soc., 106 (1984) 811.
159. J.W. Brady, *Adv. Biophys. Chem.*, 1 (1990) 155.
160. A.D. French and J.W. Brady (eds.), *Computer Modeling of Carbohydrate Molecules*. ACS Symposium Series 430, American Chemical Society, Washington, DC, 1990.
161. B. Meyer, *Topics Current Chem.*, 154 (1990) 141.
162. S. Pérez and M.M. Delage, *Carbohydr. Res.*, 212 (1991) 253.
163. S.W. Homans and M. Forster, *Glycobiology* 2 (1992) 143.
164. J.W. Brady and R.K. Schmidt, *J. Phys. Chem.*, 97 (1993) 958.
165. C.J. Cramer and D.G. Truhlar, *J. Am. Chem. Soc.*, 115 (1993) 5745.
166. A.J. Duben, M. Hricovíni and I. Tvaroska, *Carbohydr. Res.*, 247 (1993) 71.
167. A.D. French and M.K. Dowd, *Theochem: J. Mol. Structure*, 105 (1993) 183.
168. P.D.J. Grootenhuis and C.A.G. Haasnoot, *Mol. Simulation*, 10 (1993) 75.
169. B.J. Hardy and A. Sarko, *J. Comp. Chem.*, 14 (1993) 831.
170. L.M.J. Kroonbatenburg, J. Kroon, B.R. Leeftang and J.F.G. Vliegthart, *Carbohydr. Res.*, 245 (1993) 21.
171. T. Peters, B. Meyer, R. Stuikeprill, R. Somorjai and J.-R. Brisson, *Carbohydr. Res.*, 238 (1993) 49.
172. M.J. Forster and B. Mulloy, *J. Comput. Chem.*, 15 (1994) 155.
173. T.L. James, *Curr. Opin. Struct. Biol.*, 4 (1994) 275.
174. S.W. Homans, *Glycobiology* 3 (1993) 551.
175. L. Poppe and H. van Halbeek, *J. Am. Chem. Soc.*, 114 (1992) 1092.
176. L.E. Lerner, in: *Synthetic Carbohydrates* (P. Ková, ed.), ACS Symposium Series, American Chemical Society, Washington, DC, 1994.
177. D.A. Cumming and J.P. Carver, *Biochemistry*, 26 (1987) 6664.
178. M. Hricovíni, R.N. Shah and J.P. Carver, *Biochemistry*, 31 (1992) 10018.
179. I. Braccini, V. Michon, C. Hervé du Penhoat, A. Imberty and S. Pérez, *Int. J. Biol. Macromol.*, 15 (1993) 52.
180. T. Peters and T. Weimar, *J. Biomolec. NMR*, 4 (1994) 97.
181. G. Lipari and A. Szabo, *J. Am. Chem. Soc.*, 104 (1982) 4546.
182. G. Lipari and A. Szabo, *J. Am. Chem. Soc.*, 104 (1982) 4559.
183. Y. Aubin and J.H. Prestegard, *Biochemistry*, 32 (1993) 3422.
184. Y. Aubin, Y. Ito, J.C. Paulson and J.H. Prestegard, *Biochemistry*, 32 (1993) 13405.
185. L. Poppe, H. van Halbeek, D. Acquotti and S. Sonnino, *Biophys. J.*, 66 (1994) 1642.
186. K.R. Barber, K.S. Hamilton, A.C. Rigby and C.W.M. Grant, *Biochim. Biophys. Acta*, 1190 (1994) 376.
187. A. Schanck, M.P. Mingeot-Leclercq, P.M. Tulkens, D. Carrier, I.C. Smith and H.C. Jarrell, *Chem. Phys. Lipids*, 62 (1992) 153.
188. P.O. Eriksson, L. Rilfors, A. Wieslander, A. Lundberg, G. Lindblom, *Biochemistry*, 30 (1991) 4916.
189. D.B. Fenske, K. Hamilton, H.C. Jarrell, E. Florio, K.R. Barber and C.W.

- Grant, *Biochemistry*, 30 (1991) 4503.
190. I.C. Smith, J. Baenziger, M. Auger and H.C. Jarrell, *Prog. Clin. Biol. Res.*, 292 (1989) 13.
  191. J.T. Davis, S. Hirani, C. Bartlett and B.R. Reid, *J. Biol. Chem.*, 269 (1994) 3331.
  192. G. Weisshaar, J. Hiyama and A.G.C. Renwick, *Glycobiology*, 1 (1991) 393.
  193. J.M. Withka, D.F. Wyss, G. Wagner and A.R.N. Arulanandam, *Structure* 1 (1993) 69.
  194. S.J. Glaser and G.P. Drobny, *Chem. Phys. Lett.*, 164 (1989) 456.
  195. S.J. Glaser and G.P. Drobny, *Chem. Phys. Lett.*, 184 (1991) 553.
  196. M. Gilleron, J. Vercauteren and G. Puzo, *J. Biol. Chem.*, 268 (1993) 3168.
  197. M. Rivière, S. Augé, J. Vercauteren, E. Wisingerová and G. Puzo, *Eur. J. Biochem.*, 214 (1993) 395.
  198. D.T. Connolly, R.R. Townsend, K. Kawaguchi, W.R. Bell and Y.C. Lee, *J. Biol. Chem.*, 257 (1982) 939.
  199. S. Sabesan, J.Ø. Duus, S. Neira, P. Domaille, S. Kelm, J.C. Paulson and K. Bock, *J. Am. Chem. Soc.*, 114 (1992) 8363.
  200. S.A. DeFrees, F.C.A. Gaeta, Y.-C. Lin, Y. Ichikawa and C.-H. Wong, *J. Am. Chem. Soc.*, 115 (1993) 7549.
  201. H.C. Jarrell and I.C. Smith, *Basic Life Sci.*, 56 (1990) 303.
  202. G.M. Clore and A.M. Gronenborn, *J. Magn. Reson.*, 48 (1982) 402.
  203. G.M. Clore and A.M. Gronenborn, *J. Magn. Reson.*, 53 (1983) 423.
  204. A.P. Campbell and B.D. Sykes, *J. Magn. Reson.*, 93 (1991) 77.
  205. V.L. Bevilacqua, D.S. Thomson and J.H. Prestegard, *Biochemistry*, 29 (1990) 5529.
  206. V.L. Bevilacqua, Y. Kim and J.H. Prestegard, *Biochemistry*, 31 (1992) 9339.
  207. C.P.J. Glaudemans, L. Lerner, G.D. Daves, Jr., P. Ková, R. Venable and A. Bax, *Biochemistry*, 29 (1990) 10906.
  208. T. Weimar and T. Peters, *Angew. Chem.-Int. Ed. Engl.*, 33 (1994) 88.
  209. J.P. Carver, *Pure Appl. Chem.*, 65 (1993) 763.
  210. K. Bock, J.Ø. Duus, O. Hindsgaul and I. Lindh, *Carbohydr. Res.*, 228 (1992) 1-20.
  211. I. Lindh and O. Hindsgaul, *J. Am. Chem. Soc.*, 113 (1991) 216.
  212. R. Barker and T.E. Walker, *Meth. Carbohydr. Chem.*, 8 (1980) 151.
  213. J. Wu and A.S. Serianni, *Carbohydr. Res.*, 226 (1992) 209.
  214. A.S. Serianni and P.B. Bondo, *J. Biomol. Struct. Dynamics*, 11 (1994) 1133.
  215. A.S. Serianni, *Stable Isotopically-Labeled Carbohydrates and Nucleosides: Synthesis and Applications in Chemistry and Biology*, in: *Stable Isotope Applications in Biomolecular Structure and Mechanisms*, J. Trewhella, T.A. Cross and C.J. Unketer, eds., Los Alamos National Laboratory, Los Alamos, NM, 1994.
  216. D.G. Davis, *J. Magn. Reson. Series B*, 101 (1993) 229.

217. S.W. Fesik, R.T. Gampe, Jr., H.L. Eaton, G. Gemmecker, E.T. Olejniczak, P. Neri, T.F. Holzman, D.A. Egan, R. Edalji, R. Simmer, R. Helfrich, J. Hochlowski and M. Jackson, *Biochemistry*, 30 (1991) 6574.
218. V.L. Hsu and I.M. Armitage, *Biochemistry*, 31 (1992) 12778.
219. S.I. Yamakage, T.V. Maltseva, F.P. Nilson, A. Foldesi and J. Chattopadhyaya, *Nucl. Acids Res.*, 21 (1993) 5005.

## Chapter 8

# Solid State NMR and its Applications to Biomedical Research

ALEXANDRA SIMMONS, SUSANTA K. SARKAR  
and LYNN W. JELINSKI

## 1. INTRODUCTION

Multidimensional NMR has proven to be a valuable tool for establishing the structure and dynamics of biological molecules in solution. Solution NMR is frequently combined with X-ray diffraction and molecular mechanics calculations to determine the extent to which solution properties (e.g. conformation and dynamics) are retained in the solid state. However, there are many biological molecules that cannot be prepared as single crystals. In addition, the presence of solvent may have a profound effect on the structure and dynamics of individual molecules, and change the nature of intermolecular interactions. Solution techniques are not adequate for the examination of molecules or assemblies of molecules in which motion is restricted, so that isotropic tumbling fast on the NMR time scale does not occur.

Techniques applied to study solid samples of synthetic polymers and inorganic glasses can also be applied to questions of biological interest. This chapter outlines the differences between nuclear interactions in solution and in the solid state, then introduces the techniques commonly used to obtain useful spectra from solid samples. Next, a number of applications are presented in which information uniquely available from solid-state NMR has increased our understanding of a biological material or process. The wide scope of this chapter precludes a complete listing of all related research. Therefore, only selected references will be given, and the nuclei considered will be limited to  $^1\text{H}$ ,  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ , and  $^{31}\text{P}$ .

## 2. FUNDAMENTALS OF SOLID-STATE NMR

### 2.1. Nuclear interactions in the solid state

The nuclear interactions discussed here will be described from a physical point of view, rather than a rigorous quantum-mechanical one. A more detailed description of these phenomena can be obtained in Slichter [1] or Mehring [2], while Fyfe is a general solid-state NMR text aimed at chemists [3]. The types of interactions important in NMR experiments on spin-1/2 nuclei are, in order of decreasing strength, Zeeman, dipolar, chemical shift, and scalar. The only nucleus with spin greater than 1/2 considered here is  $^2\text{H}$ . A description of  $^2\text{H}$  spectroscopy will require introduction of another type of nuclear interaction, namely quadrupolar coupling.

#### *Zeeman*

The Zeeman interaction of the spin with the static magnetic field removes the degeneracy of the nuclear energy levels. The energy difference between levels is proportional to the magnetic field: for  $^{13}\text{C}$  nuclei in a 4.7 Tesla magnetic field (which we will use as a reference nucleus), the magnitude of the Zeeman interaction is 50.3 MHz. The size of the Zeeman interaction is the same in solution or in the solid state, and it is about 100 times larger than any of the other interactions we will consider.

#### *Dipolar*

The dipolar interaction brings about direct magnetic coupling of nuclei through space. It is not field-dependent. For a  $^{13}\text{C}$  nucleus with a bonded proton, one can write the local field as

$$B_{\text{loc}} = \pm \frac{h}{4\pi} \gamma_{\text{H}} \frac{(3\cos^2\theta_{\text{C-H}} - 1)}{r_{\text{C-H}}^3}$$

where  $\gamma_{\text{H}}$  is the proton magnetogyric ratio,  $r_{\text{C-H}}$  is the carbon-proton internuclear distance, and  $\theta_{\text{C-H}}$  is the angle the C-H vector makes with the static magnetic field. The dipole-dipole interaction is averaged to zero by rapid tumbling in solution. By contrast, in solids the dipolar interaction is large (about 50 kHz for  $^{13}\text{C}$  at 4.7 T), and its magnitude depends on the orientation of the internuclear vector with respect to the static magnetic field. For a  $^{13}\text{C}$ - $^1\text{H}$  pair in a single crystal, a splitting of



the signal into two components takes place. In glassy or microcrystalline samples, the summation over all  $\theta_{C-H}$  and  $r_{C-H}$  values results in substantial line broadening by dipolar interactions.

### *Chemical-shift anisotropy (CSA)*

Just as in solution, the electronic screening of a nucleus in the solid state affects its resonance frequency or chemical shift. It is not readily apparent in solution that chemical shift, like the dipolar interaction, is a tensor quantity which is directional; that is, it depends on the orientation of the molecule with respect to the magnetic field. The isotropic chemical shift observed for a rapidly tumbling molecule in solution is one third of the trace of the chemical shift tensor,

$$\sigma_{\text{isotropic}} = 1/3 (\sigma_{11} + \sigma_{22} + \sigma_{33})$$

where  $\sigma_{11}$ ,  $\sigma_{22}$ , and  $\sigma_{33}$  are the values of the chemical-shift tensor in the principal axis system. In the absence of motional averaging in a solid, each magnetically inequivalent site gives rise to a powder lineshape that reflects the site symmetry and the principal components of the chemical-shift tensor. The  $^{13}\text{C}$  CSA in a solid may be as large as 200 ppm.

### *Scalar*

Scalar or J coupling of two nuclei takes place through the bonding electrons. It cannot be averaged by rapid molecular tumbling, so its magnitude, about 200 Hz, is similar in solution and in the solid state. Although useful for structural assignments in solution, scalar couplings are rarely observed in the solid state because they are obscured by the much larger interactions described above.

In the solid state, internuclear interactions may cause such severe broadening that a single broad peak spanning the entire chemical-shift range is observed. Techniques for obtaining high-resolution spectra from solid samples will be presented next. However, because internuclear interactions are orientation- and distance-dependent and sensitive to motion, they can be a source of structural and dynamic information. Techniques to selectively reintroduce certain interactions under high-resolution conditions have recently been developed, and they will be covered briefly.

## 2.2. High-resolution spectra of dilute spin-1/2 nuclei ( $^{13}\text{C}$ and $^{15}\text{N}$ )

Twenty years have not yet passed since Schaefer and Stejskal, using home-built equipment, combined magic-angle spinning (MAS), dipolar decoupling (DD) and cross-polarization (CP) to obtain high-resolution  $^{13}\text{C}$  spectra of cellulose and collagen, two biological macromolecules [4]. Due to their wide applicability and relatively easy implementation, the techniques presented here are now routinely applied using commercial spectrometers and probes. Each of these techniques, magic-angle spinning, dipolar decoupling and cross-polarization, will be described in turn.

### *Magic-angle spinning [5,6]*

The  $^{13}\text{C}$  CSA at 4.7 T is at most 5–10 kHz. Air-driven rotors have been designed to spin the sample at frequencies greater than these. In that case, the direction cosines of the CSA become time-dependent in the rotor period, and the time average becomes

$$\sigma = 1/2 \sin^2\theta(\sigma_{11} + \sigma_{22} + \sigma_{33}) + 1/2(3 \cos^2\theta - 1) \text{ (functions of direction cosines)}$$

where  $\theta$  is the angle between the magnetic field and the axis of rotation. When  $\theta$  is  $54.7^\circ$ , then the above equation reduces to the isotropic chemical shift, hence this is known as the “magic angle”. Magic-angle spinning (MAS) at frequencies greater than those of the CSA results in a single peak for each inequivalent nucleus at its isotropic chemical shift value. Slower spinning yields a peak at the isotropic chemical shift flanked by sidebands separated by the spinning frequency, as shown in Fig. 1 for glycine [7]. The sidebands can be used to provide information about the size of the individual chemical shift tensors, and therefore molecular motion in the sample. The necessity of performing MAS results in one of the major differences between solution and solid-state NMR: in solution, experiments are carried out at the maximum possible field in order to improve signal-to-noise and to separate scalar couplings. Because the CSA scales with the magnetic field, an increase in field requires an increase in spinning speed. MAS is routinely performed at 5 kHz, so  $^{13}\text{C}$  CSA can be almost completely eliminated at 4.7 Tesla. Higher-speed rotors are under development. Furthermore, pulse sequences such as total suppression of sidebands (TOSS) can be used to eliminate spinning sidebands if they cause problems in spectral interpretation. Consequently, obtaining high-resolution spectra at high field strengths is not nearly as difficult as once thought.

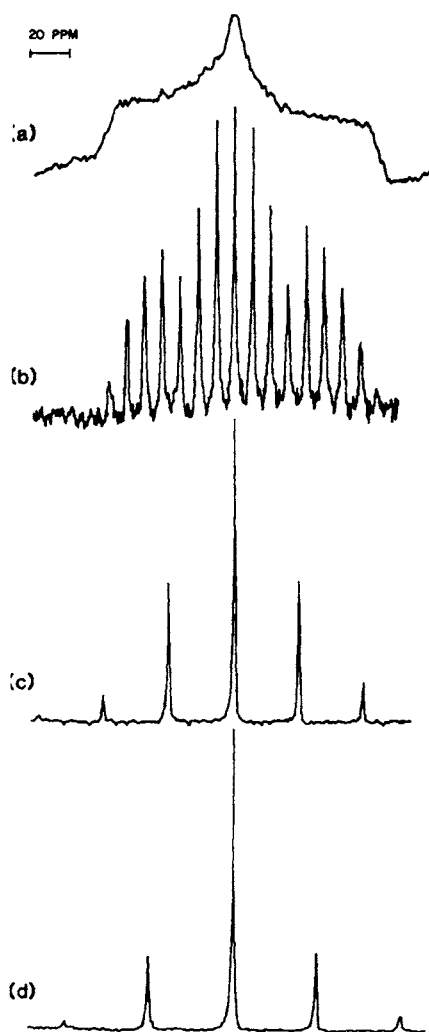


Fig. 1.  $^{13}\text{C}$  NMR spectra (50.3 MHz) of the carboxyl carbon of glycine. (a) Static powder pattern; (b) with magic-angle spinning at 0.5 kHz; (c) 1.6 kHz; (d) 2.1 kHz.

#### *Dipolar decoupling [4]*

We have seen that  $^{13}\text{C}$ – $^1\text{H}$  dipolar interactions may be around 50 kHz. Therefore, they cannot be removed by MAS at presently attainable speeds. In analogy to scalar decoupling in solution NMR, rapid inver-

sion of the proton spins relative to carbon spins by irradiation with a strong rf field at the  $^1\text{H}$  Larmor frequency reduces the dipolar coupling to zero. Whereas 5 Watts of power is usually sufficient to remove scalar coupling in solution, dipolar decoupling (DD) in solids may require 1 kW. Dipolar decoupling also averages the scalar interactions to zero. Since  $^{13}\text{C}$  is a dilute nucleus, and dipolar coupling is proportional to  $1/r^3$ , homonuclear  $^{13}\text{C}$ - $^{13}\text{C}$  coupling is usually not a concern.

### Cross-polarization [8]

MAS and DD provide sufficient line-narrowing for high-resolution spectra, which give chemical and structural information, to be obtained from dilute nuclei in solids. However, the problem of low sensitivity remains. The low sensitivity arises from two factors. One is the low natural abundance of  $^{13}\text{C}$ . The other is the relatively long  $^{13}\text{C}$  relaxation times in solids. Cross-polarization (CP) from  $^1\text{H}$  to  $^{13}\text{C}$  is a method that circumvents these two factors and is used for signal enhancement. The CP pulse sequence is shown in Fig. 2. A  $90^\circ$  pulse is applied to the protons to bring them onto the y axis of the rotating frame. A  $90^\circ$  phase shift of the  $^1\text{H}$   $B_1$  field spin-locks the protons along y, where they precess at  $\omega_{\text{H}} = \gamma_{\text{H}}B_{1\text{H}}$ . During the spin-lock time,  $^{13}\text{C}$ - $^1\text{H}$  contact is established by turning on the  $^{13}\text{C}$  field such that

$$\gamma_{\text{C}}B_{1\text{C}} = \gamma_{\text{H}}B_{1\text{H}}$$

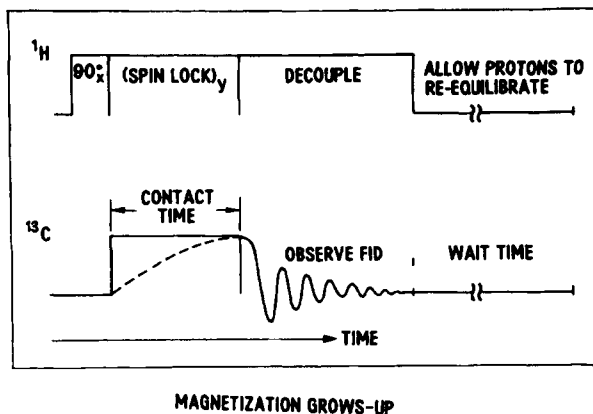


Fig. 2. Cross-polarization pulse sequence.

This is known as the Hartmann–Hahn condition [9]. This condition forces an energy overlap between  $^{13}\text{C}$  and  $^1\text{H}$  levels in the rotating frame, even though none exists in the laboratory frame. Because their  $z$  components have the same time dependence, magnetization transfer from  $^1\text{H}$  to  $^{13}\text{C}$  can occur via mutual spin flip-flops. The  $^{13}\text{C}$  field is then turned off, and acquisition takes place with dipolar decoupling. The enhancement in signal intensity can be as large as  $\gamma_{\text{H}}/\gamma_{\text{C}}$ , a factor of about 4.

Homonuclear dipolar coupling between protons allows spin diffusion to occur. This reduces the spin-lattice relaxation time ( $T_1$ ) of all the protons to that of the fastest-relaxing species, usually a  $-\text{CH}_3$  group. Since  $^{13}\text{C}$  nuclei are far apart, homonuclear dipolar coupling is weak; restricted motion usually results in long  $^{13}\text{C}$   $T_1$  values, even at the fastest-moving sites. Without CP, the pulse repetition rate is limited by the  $^{13}\text{C}$   $T_1$ s. With CP, the magnetization is generated by the  $^1\text{H}$  bath, so it is only necessary to wait for the protons to relax before pulsing again. This can result in a many-fold increase in repetition rate, which adds to the signal enhancement obtained by magnetization transfer.

The combination of MAS, CP and DD brings about a spectacular improvement in spectral resolution not possible with any one of these techniques. The spectrum of crystalline sucrose with and without line-narrowing techniques is shown in Figure 3 along with its solution spectrum [10]. The resolution attainable in solution is still better than that of high-resolution solid-state spectra. However, solid samples often show additional peak splittings or conformation-dependent chemical shifts which indicate important differences between conformation or secondary structure in solution and the solid state.

Modifications of the CP pulse sequence shown above can be used to obtain high-resolution spectra from which spin-lattice relaxation times in the laboratory frame ( $T_1$ ) or in the rotating frame ( $T_{1\rho}$ ) can be calculated. Since the signal is generated by CP, these parameters can be determined for the abundant nuclei (usually  $^1\text{H}$ ) or for the dilute nuclei. Each of these relaxation times is sensitive to motion at a certain frequency. For example,  $^{13}\text{C}$   $T_1$  measurements are sensitive to motion at frequencies from  $10^5$  to  $10^{10}$  Hz and above, whereas  $^{13}\text{C}$   $T_{1\rho}$  responds to motions between  $10^1$  and  $10^3$  Hz. Therefore, like CSA or lineshape, relaxation can be used to study molecular dynamics while the high-resolution spectrum itself gives structural information.

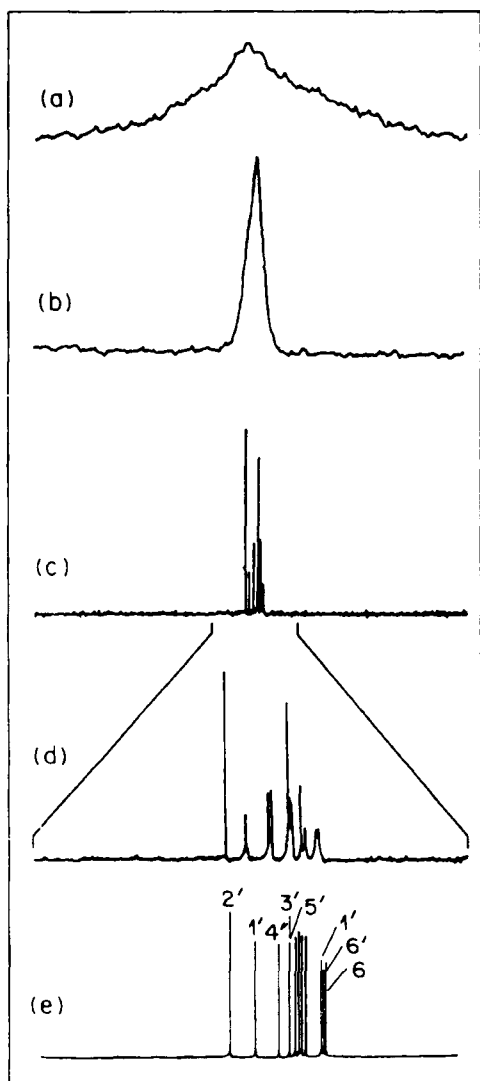


Fig. 3.  $^{13}\text{C}$  NMR spectra (50.3 MHz) of crystalline sucrose: (a) no MAS, no DD; (b) no MAS, with DD; (c) MAS, DD, CP; (d) expansion of (c); (e) in aqueous solution. MAS, DD, and CP can be applied to other dilute nuclei. In fact for  $^{15}\text{N}$ ,  $\gamma_{\text{H}}/\gamma_{\text{N}}$  is nearly 10, resulting in even greater enhancement. The combination of CP, MAS, and DD has made the acquisition of high-resolution spectra of dilute nuclei in the solid state routine.

### 2.3. *High-resolution solid-state spectra of abundant nuclei ( $^1\text{H}$ and $^{31}\text{P}$ )*

Although  $^{31}\text{P}$  is the only naturally-occurring isotope of phosphorus, it can often be considered to be a rare nucleus. The concentration of  $^{31}\text{P}$  nuclei in an organic sample may be much lower than that of  $^{13}\text{C}$  or  $^1\text{H}$ . Even though all the P nuclei are the active  $^{31}\text{P}$  isotope, individual nuclei may be quite far from each other. In such a sample, homonuclear  $^{31}\text{P}$ - $^{31}\text{P}$  interactions are weak, and the techniques described above (CP, MAS, DD) can be applied.

Before techniques to obtain high-resolution spectra from dilute nuclei were developed, featureless broadband spectra of abundant nuclei like  $^1\text{H}$  were used to study solids. By performing mathematical analyses including lineshape modeling and second moment analysis, some information, such as motional rates, could be extracted. It is clearly desirable to obtain high-resolution spectra of abundant nuclei. For  $^1\text{H}$  or abundant  $^{31}\text{P}$  some modifications to those procedures described for dilute nuclei are required. MAS will provide some line narrowing, but will not be fast enough to remove the homonuclear dipolar interactions. Dipolar decoupling cannot be used, as one cannot simultaneously irradiate and observe the same nucleus.

#### *Multiple-pulse line narrowing*

Several pulse sequences have been developed to remove the homonuclear dipolar interaction. Multiple-pulse line narrowing works because a series of pulses of appropriate lengths applied with suitable delays can reorient the spins in predictable ways. The aim of the sequences is to average spin vectors in the rotating frame to the magic angle with respect to the applied field. One of the first and most widely used pulse sequences, known as WAHUHA [11], consists of a series of 4 phase-shifted  $90^\circ$  pulses separated by delays of  $t_d$  or  $2t_d$ . More sophisticated sequences with an increasing number of pulses and delays continue to be developed. When applied together with MAS, these sequences comprise a group of techniques for obtaining high-resolution spectra of abundant nuclei known as CRAMPS (combined rotation and multiple pulse spectroscopy).

In spite of improvements in resolution,  $^1\text{H}$  solid-state spectra are still much broader than those in solution. Because of the small chemical-shift dispersion, further progress will be required before  $^1\text{H}$  NMR becomes the standard analytical technique for structure determination which it is in solution NMR.

## 2.4. Deuterium quadrupole echo spectroscopy

Deuterium quadrupole echo spectroscopy is not a high-resolution NMR technique. On the contrary, the lineshape is so broad that much of the signal dies away during the receiver dead time. The quadrupole echo pulse sequence (Fig. 4) is designed to refocus the magnetization so that acquisition can take place once the receiver has had time to recover. Deuterium quadrupole echo spectroscopy has found wide application in biological systems because of its sensitivity to orientation and the rate and amplitude of motion at a specific site.

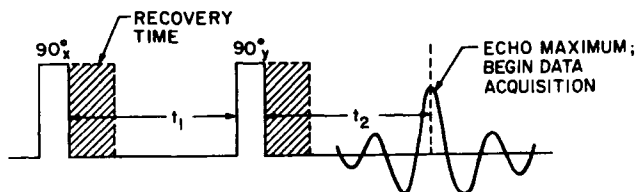


Fig. 4. Quadrupolar echo pulse sequence.

The deuterium nucleus has a spin of 1, so the Zeeman interaction yields three equally-spaced levels. As a quadrupolar nucleus, there is a non-spherical charge distribution at the nucleus. The interaction of the quadrupole moment with the nuclear electric field gradient perturbs the Zeeman splitting more than any type of interaction mentioned above, by up to 250 kHz. The size of the quadrupolar splitting,  $\Delta\nu_{\text{quad}}$ , is related to the angle the C–D bond makes with the magnetic field ( $\theta_{\text{C-D}}$ ):

$$\Delta\nu_{\text{quad}} = 3/4 \frac{e^2qQ}{h} (3\cos^2\theta_{\text{C-D}} - 1)$$

where  $e$  is the charge of the electron and  $e^2qQ/h$  is the quadrupolar coupling constant. The  $(3\cos^2\theta - 1)$  orientation dependence of the quadrupolar splitting is identical to that given above for the CSA or dipolar broadening. In the same way, a powder lineshape is built up from signals from C–D bonds at all orientations, according to their probabilities as shown in Fig. 5a.

A static C–D bond gives rise to the expected Pake powder pattern with a quadrupolar splitting of 128 kHz. If molecular motion occurs that



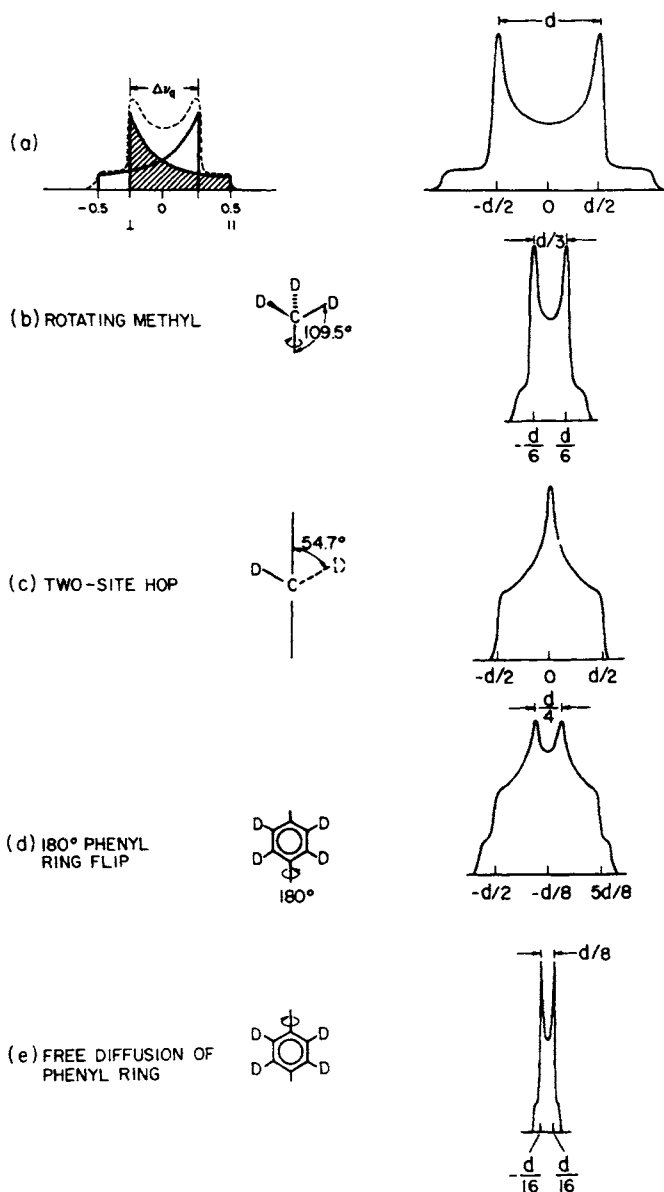


Fig. 5. Theoretical  $^2\text{H}$  lineshapes for several types of rapid ( $>10^7$  Hz) anisotropic motions.

is rapid on the NMR time scale (greater than  $10^7$  Hz), the quadrupolar coupling and lineshape are affected in a predictable way. Figure 5 shows deuterium lineshapes for specific types of motions. For example, in a rapidly-rotating methyl group the C–D bonds form an angle of  $109.5^\circ$  with the rotation axis, and the lineshape collapses by a factor of  $1/2(3\cos^2\theta - 1)$  to  $1/3$  of the static splitting. It can be seen that lineshape changes brought about by different types of motions are easily distinguished.

We have seen that deuterium lineshapes are sensitive to the amplitude of motion at a given site. Motions with intermediate frequencies on the deuterium NMR time scale ( $10^3$ – $10^7$  Hz) cause fluctuation of the NMR transition between two values in a manner analogous to chemical exchange in high-resolution spectra. Deuterium lineshapes calculated for such motions can provide rate constants to plot against inverse temperature which give Arrhenius activation energies.

The deuterium lineshape is sensitive to motion at intermediate rates, and gives information about the amplitude of fast motions. Insight into the rates of fast motions can be gained from deuterium spin-lattice relaxation times ( $T_1$ ).  $T_1$  can be determined by an inversion-recovery sequence like that used in solution, followed by detection using a quadrupolar echo.

In most systems, the quadrupolar interaction completely dominates the splitting and the relaxation of deuterium nuclei. Therefore, unambiguous conclusions about molecular-level structure and dynamics can be reached. The low natural abundance (0.016%) of deuterium would seem to be a disadvantage. In fact, through selective labeling, it allows investigation of a specific site without background interference.

## ***2.5. Distance measurements with solid-state NMR***

We have already mentioned that the magnitude of dipolar couplings are strongly dependent on the internuclear distance. However, because of their broadening effect, dipolar couplings are usually removed to acquire high-resolution spectra. Techniques have been developed to reintroduce dipolar coupling between isolated pairs of nuclei while obtaining high-resolution spectra. These experiments allow very precise measurement of internuclear distances of up to 5 Å in amorphous or microcrystalline samples. This means that structural details formerly available only by X-ray diffraction can be calculated even in samples not amenable to X-ray analysis. Insofar as these techniques use a through-space interaction to determine internuclear distance,

they are analogous to Nuclear Overhauser Enhancement (NOE) experiments in solution.

#### *Rotational resonance [12]*

Griffin has developed a method to measure homonuclear couplings in amorphous solids. Dipolar couplings between two spin-1/2 species are averaged by sample spinning. However, at rotational resonance,

$$\Delta = n\omega_r$$

where  $\Delta$  is the difference in isotropic chemical shifts of the two nuclei and  $\omega_r$  is the spinning speed, a reduced form of the homonuclear dipolar coupling reappears. The reintroduction of this coupling between an isolated spin pair has a theoretically calculable result on the spectrum. A lineshape simulation can be used to calculate the magnitude of the dipolar interaction and determine through-space homonuclear distances.

#### *Rotational-echo double resonance (REDOR) [13]*

This technique uses MAS and measures the heteronuclear dipolar coupling between isolated pairs of labeled nuclei to determine interatomic distances of several Å. For example, in a solid with  $^{13}\text{C}$ – $^{15}\text{N}$  dipolar coupling, the  $^{13}\text{C}$  rotational echoes that form each rotor period after the  $^1\text{H}$ – $^{13}\text{C}$  CP transfer can be prevented from reaching full intensity by inserting two  $^{15}\text{N}$   $\pi$ -pulses per rotor period.

The difference between a  $^{13}\text{C}$  spectrum obtained this way and one with no  $^{15}\text{N}$   $\pi$ -pulses measures the  $^{13}\text{C}$ – $^{15}\text{N}$  coupling. The magnitude of the coupling allows the distance between the two nuclei to be calculated explicitly, and the incorporation of labeled nuclei allows the exact site to be chosen.

## **2.6. Summary**

Solid-state NMR is now recognized as a method to obtain site-specific information about molecular-level structure and dynamics of large and small molecules and assemblies of molecules in the solids. Experiments can be designed to be sensitive to particular frequencies and amplitudes of molecular motion. In many cases, no other method is able to supply this information. With the advent of techniques to measure internuclear distances precisely, new insights into molecular structure are being gained. The rest of this chapter provides selected examples of the application of these techniques to questions of biomedical interest.

### 3. APPLICATIONS OF SOLID-STATE NMR TO QUESTIONS OF BIOMEDICAL INTEREST

In the preceding sections, we saw how developments in the theory and practice of NMR in solids make it possible to obtain high-resolution spectra of solid samples. This includes solid crystalline or amorphous samples or compounds in solution that are part of large aggregates and are immobile or reorient slowly on the NMR time scale. Therefore, a variety of biological samples including large ordered systems such as membranes or nucleic acids can be characterized with atomic resolution. On the other hand, wideline spectra contain a great deal of information about nuclear interactions and molecular dynamics. Spectra of oriented samples provide information about the orientation of the principal axis system of the molecule or functional group with respect to the external magnetic field. We shall see that many structural studies take advantage of the spectral simplification brought about by sample orientation. Single lines characterized by the chemical shift value or doublets which yield dipolar splittings can be used to orient nuclei relative to the field and to each other. A number of independent constraints on nuclear positions or bond angles can be combined to build up complete structures for compounds in the solid state.

#### 3.1. *Bone*

Because they are so small, the crystallites in the mineral phase of bone generate only a few diffuse reflections in X-ray diffraction experiments. Interest in determining the chemical and structural nature of  $\text{CaPO}_4$  in bone led to the application of solid-state  $^{31}\text{P}$  MAS NMR to mineral deposits in bone from several sources at different ages. For example, in chicken bone [14], three types of pulse sequences were used: a Bloch decay sequence, a CP sequence and a sequence in which removal of  $^1\text{H}$  decoupling for a short time before acquisition leads to dipolar suppression of the protonated brushite (BRU)-like component. The smaller CSA of hydroxyapatite (HA) relative to BRU means that the central peak of spectra taken with the Bloch decay sequence is dominated by HA. Protons bonded to  $^{31}\text{P}$  in BRU enhance its intensity in the CP spectrum. The use of these three sequences and a comparison to the spectrum of a variety of crystalline and non-crystalline reference standards showed that calcium phosphate is present as a type B HA containing 5–10% of  $\text{CO}_3^{2-}$  and 5–10% of  $\text{HPO}_4^{2-}$  in a BRU-like configura-

tion. The fraction of  $\text{HPO}_4^{2-}$  was shown to decrease with the age of the bone. The difference in the  $^{31}\text{P}$   $T_1$  of calcium-deficient HA and pure HA or  $\beta$ -tricalcium phosphate led to the classification of rabbit bone as calcium-deficient HA and allowed the analysis of newly-deposited bone in  $\beta$ -tricalcium phosphate implants [15]. A  $^{31}\text{P}$  MAS study of embryonic chick bone fractions detected only a poorly crystalline HA phase. Although some of the phosphate was seen to be  $\text{HPO}_4^{2-}$  in a BRU-like configuration incorporated into the HA lattice, no evidence of a precursor phase of a different structure or composition was found [16]. The poor X-ray diffraction properties of bone minerals and the necessity of studying them in the solid state means that solid-state NMR is the only technique able to provide us with this information.

### ***3.2. Lipid bilayers and membranes***

The application of solution NMR techniques to lipid bilayers or membranes usually results in the observation of broad featureless spectra. This is because such samples are solid or semisolid liquid crystals in which restricted anisotropic motion leads to incomplete averaging of the nuclear interactions. Sonication reduces aggregate size and improves resolution in solution, but may change the bilayer properties. Phospholipids undergo several thermal phase transitions. At low temperatures or after annealing, a highly ordered subgel ( $L_c$ ) phase may exist; heating leads to the formation of a gel ( $L_\beta$ ) phase, which may be followed by an intermediate ripple ( $P_\beta$ ) phase, and finally a liquid-crystalline ( $L_\alpha$ ) phase. The gel ( $L_\beta$ ) to liquid-crystal ( $L_\alpha$ ) transition is the main transition, with the transition temperature denoted as  $T_c$ . Thermal analysis can detect the phase changes but gives no information about molecular-level changes in structure or dynamics. Many hundred papers have appeared which use solid-state  $^1\text{H}$ ,  $^2\text{H}$ ,  $^{13}\text{C}$  or  $^{31}\text{P}$  NMR to determine the molecular level orientation and dynamics of membranes and membrane models, to characterize phase transitions or to observe interactions between lipids and other compounds. What follows is a selection of these papers which illustrates the utility of some of these techniques.

As early as 1974, lipid bilayers were analyzed by  $^2\text{H}$  solid-state NMR to obtain quantitative information on ordering and motional anisotropy at various parts of the molecules. Quadrupole splittings were measured for hydrated dipalmitoyl phosphatidylcholine (DPPC) in the  $L_\alpha$  phase specifically deuterated at 9 different acyl carbons [17]. Segmental order

parameters were found to be constant for the first nine segments, with quadrupole splitting increasing, and order decreasing, towards the chain's methyl terminus. Heating was observed to decrease the order parameter. The data indicated the presence of some gauche conformers, suggesting that the ordered bilayer contained disordered hydrocarbon chains. Determination of the order parameter of the labeled segment allowed the average chain geometry and therefore the number of gauche conformers to be estimated, so that the thickness of the hydrocarbon region of the bilayer was estimated to be 35 Å. A  $^2\text{H}$  study of hydrated chain-perdeuterated DPPC followed in which it was shown that the spectrum could be simulated by the superposition of a number of quadrupolar powder patterns [18]. Well below  $T_c$  a static pattern was observed, but even 20°C below  $T_c$  the C–D bond order parameters and the lineshape suggested that rapid reorientation was occurring. At  $T_c$  there was a sudden decrease in the first and second moments of the spectra. A continuous slow decrease in the moments on further heating reflected an overall decrease in chain order due to increased motional averaging of the quadrupolar interaction.

$^{31}\text{P}$  is an obvious choice for the study of phospholipids as it occurs naturally in these compounds and is a high-sensitivity nucleus; it is a convenient probe of headgroup motions. In addition, model shift tensors to aid in spectral interpretation are available. MAS at 2.6 kHz narrows the 60 ppm wide  $^{31}\text{P}$  CSA pattern of hydrated DPPC to a 95 Hz wide central peak and two sets of spinning sidebands [19]. The first high-resolution proton-decoupled  $^{13}\text{C}$  MAS spectra of DPPC and dimyristoylphosphatidylcholine (DMPC) were obtained in 1978 [19]. There was considerable overlap of CSA patterns in the static sample, although the  $-\text{N}(\text{CH}_3)_3$ , glycerol backbone, headgroup  $-\text{CH}_2-$  signals were discernible. Under MAS, one observes all the lines seen in a  $\text{CHCl}_3$  solution of DPPC, and can assign them by reference to solution spectra.

Analysis of carbonyl CSAs is useful for studying lateral diffusion in bilayers and motion of lipid molecules about their long axis. DPPC, DMPC and distearylphosphatidylcholine (DSPC) were  $^{13}\text{C}$  labeled at the *sn*-2 carbonyl, and the temperature dependence of the CSA was monitored [20]. The 100 ppm-wide axially symmetric CSA powder pattern observed in the gel phase became an isotropic line at  $T_c$ . At temperatures slightly below  $T_c$ , a superposition of the  $L_\beta$  and  $L_\alpha$  patterns was seen, indicating an intermediate ( $P_\beta$ ) phase with some attributes of both the gel and liquid-crystalline phases. Fast axial motion in all phases and a conformational change at the carbonyl upon passing

through the main phase transition were consistent with the data. In the  $P_\beta$  phase, both conformations are present, and exchange occurs, with the rate constant for exchange related to lateral diffusion along the bilayer.

Phosphatidylethanolamines (PE) are another class of lipids in biological membranes. DipalmitoylPE (DPPE) was labeled with  $^{13}\text{C}$  at the *sn*-2 carbonyl and  $^2\text{H}$  labeled at the 4,4- 8,8- or 12,12- acyl carbons [21]. Both  $^{13}\text{C}$  and  $^2\text{H}$  spectra of the acyl chains indicated diffusion about the long axis at rates of  $10^5$ – $10^6$  Hz even in the gel phase.  $^{13}\text{C}$  CSA spectra implied that the main transition was accompanied by the same conformational change at the carbonyl as in phosphatidylcholines (PC). No intermediate phase was seen. Sharp axially symmetric  $^2\text{H}$  spectra inferred increased disorder and fast rotational diffusion in the  $L_\alpha$  phase. Lineshape changes in the deuterated headgroup were not accompanied by spectral narrowing at the main transition. This was accounted for by faster head group motions in the  $L_\alpha$  phase, but no change in the distribution of conformations of the head group. Biological phospholipids are frequently heterogeneous in their acyl chain composition. Therefore, two mixed chain PCs, 1-myristoyl-2-palmitoylPC (MPPC) and 1-myristoyl-2-stearoylPC (MSPC) were studied by  $^{31}\text{P}$ ,  $^{13}\text{C}$ , and  $^2\text{H}$  NMR [22]. In contrast to DPPC,  $^{13}\text{C}$  and  $^{31}\text{P}$  CSA patterns showed that MSPC and MPPC go from a subgel phase in which there is no rapid axial diffusion, directly to a  $P_\beta$  phase, without passing through the  $L_\beta$  phase. However, the  $^2\text{H}$  spectra do indicate slow motion in the sub-phase, which must be due to *trans-gauche* isomerization.

All the studies cited above were carried out on fully hydrated bilayers. In nature, dehydration usually destroys lipid membranes; however, the sugar trehalose (TRE) is known to stabilize dry membranes.  $^{31}\text{P}$  NMR of the lipid headgroup and  $^2\text{H}$  NMR of the 7,7-acyl chain-labeled dry DPPC and 2:1 TRE/DPPC mixtures was carried out [23]. A thermal transition near the  $T_c$  of hydrated DPPC was observed.  $^{31}\text{P}$  CSA indicated a rigid  $\text{PO}_4$  headgroup with axial diffusion absent both below and above the new transition. The  $^2\text{H}$  spectrum of the TRE/DPPC mixture narrows by a factor of four on passing through the transition, but remains axially asymmetric excluding fast long axis diffusion, in agreement with the  $^{31}\text{P}$  data. Spectral simulation requires highly disordered acyl chains with several gauche isomers. This would require the bilayer to expand laterally on addition of TRE.  $^2\text{H}$  experiments on DPPC labeled at the choline headgroup as well as several specific acyl carbons yielded more molecular-level information about the role of TRE in

membrane stabilization [24]. The  $^2\text{H}$  spectra of the choline headgroup show more headgroup motion in the stabilized mixture than in dry DPPC. The  $^2\text{H}$  spectra also indicate that above the phase transition the acyl chains are disordered along their entire length. These data suggest that TRE interacts with the hydrophilic regions of DPPC to expand the distance between lipids in the bilayer. This stabilizes the membrane, but does not lead to a structure identical with that of the hydrated membrane, because axial diffusion is absent below the transition temperature of the TRE/DPPC mixture, whereas we saw above that it is present in hydrated DPPC.

The above studies took advantage of the orientation-dependent information available in wide-line  $^{13}\text{C}$ ,  $^{31}\text{P}$  and  $^2\text{H}$  spectra. It was mentioned that high-resolution  $^{13}\text{C}$  spectra of lipid bilayers could be obtained. In fact, MAS leads to high-resolution  $^1\text{H}$  and  $^{13}\text{C}$  spectra in membranes and lipids [25]. Resolution in  $^1\text{H}$  MAS spectra is as good as that in spectra of sonicated samples. This improvement is possible because in these lipids in the liquid-crystalline phase the intermolecular dipole-dipole interactions are averaged by fast lateral diffusion, while the intramolecular interactions are averaged by fast axial rotation. The  $^1\text{H}$  resolution is sufficient that  $T_1$ ,  $T_2$ , and  $T_{1\rho}$  values can be calculated. It may be possible to use the  $^1\text{H}$  sideband intensities to calculate the order parameter without the need for  $^2\text{H}$  labeling.  $^{13}\text{C}$  MAS spectra are superior to those observed by sonication: the resolution is equivalent, and some peaks appear which are not detectable in sonicated samples.  $^{13}\text{C}$  CP/MAS NMR can also be used to observe other compounds, for example sterols, in membranes. Two mixtures, one containing a chain-deuterated lipid, are required. The chain deuteration prevents cross-polarization to the membrane lipid and allows the other component to be observed. The resolution obtained by MAS methods allows high-resolution spectra of components of biological membranes, like human myelin, to be acquired and assigned.

After the acquisition of high-resolution  $^1\text{H}$  and  $^{13}\text{C}$  spectra of phospholipid bilayers, the next logical step was the design of a two-dimensional (2D) experiment. A 2D heteronuclear correlation spectrum of hydrated DMPC has been reported [26]. All resolvable directly bonded pairs appear in the spectrum, with 6 out of 14 hydrocarbon positions distinguishable from the main methylene  $^1\text{H}$  peak. All the headgroup resonances are resolved and are correlated to their respective  $^1\text{H}$  resonances.  $^2\text{H}$  lineshape analyses mentioned above are sensitive to motion in the  $10^4$ – $10^6$  kHz range. Transverse relaxation experiments indicate



the presence of some slower motions [27] in lipid bilayers. Two-dimensional  $^2\text{H}$  solid-state NMR of specifically deuterated samples was developed to characterize the rate and mode of these slow motions. Two-dimensional  $^2\text{H}$  exchange NMR of DPPC showed these motions to be related to the lateral diffusion of lipid molecules over the curved surface of the liposomes, with complete molecular orientational exchange on a timescale of 100 ms [28]. In  $^{31}\text{P}$  2D solid-state NMR, orientational exchange from lateral diffusion is detected by an increase in off-diagonal intensity, which correlates the  $90^\circ$  and  $0^\circ$  orientations of the membrane normal with respect to the magnetic field when the system is fully exchanged.  $^{31}\text{P}$  2D NMR was used to determine that the correlation time for lateral diffusion in the liquid-crystalline phase of DPPC, 8 ms, is two orders of magnitude shorter than in the gel phase, where it is about 900 ms [29].

One approach to spectral resolution in liquid-crystal phase PC dispersions has been to study oriented samples. Mixtures of DMPC with detergents known to promote magnetic orientation of DMPC have been used to eliminate the powder-pattern nature of the signals without totally averaging out the CSA [30]. This permitted the observations of individual  $^{13}\text{C}$  resonances and a number of  $^{13}\text{C}$ – $^{31}\text{P}$  dipolar couplings in DMPC. Estimations of the intrinsic dipolar couplings and CSAs for DMPC bilayers were made by extrapolating observed shifts and couplings to zero detergent, and the results were similar to those obtained in DMPC by other methods.

Rotational resonance has been used to determine intra- and intermolecular distances in DPPC bilayers [31]. Intramolecular magnetization rates measured between  $^{13}\text{C}$  labels at the 1-position of the *sn*-1 acyl chain and the 2-position of the *sn*-2 acyl chain yield a distance of 4.5–5.0 Å between them in the gel subphase. Intermolecular magnetization transfer rates between 1,2-[ $^{13}\text{C}$ ]DPPC and 2-[1- $^{13}\text{C}$ ] DPPC and between 1,2-[1- $^{13}\text{C}$ ] DPPC and 2-[2- $^{13}\text{C}$ ]DPPC indicate an average distance between labels of 4–5 Å. No crystal structure is presently available for DPPC. The rotational resonance intermolecular distance measurements demonstrate that the relative location and orientation of membrane lipids can be established by this technique.

### **3.3. Interactions with membrane lipids**

In addition to providing insight into molecular orientation and dynamics in lipid bilayers and membranes, solid-state NMR can be used

to determine what changes occur when interactions between membrane lipids and drugs or proteins take place. Two examples of such interactions will be presented here.

DPPC dispersions containing the bee venom peptide, melittin, have been studied using  $^2\text{H}$  NMR of  $[3', 3'\text{-}^2\text{H}_2]$  DPPC and  $[sn\text{-}2\text{-}^2\text{H}_{31}]$  DPPC and  $^{31}\text{P}$  NMR [32]. At a lipid to protein ratio of 4:1, melittin inhibits the formation of large lamellar structures in both the gel and liquid-crystal phases, although the bilayer arrangement is retained. Isotropic  $^{31}\text{P}$  and  $^2\text{H}$  lines reflect the formation of small rapidly tumbling structures. A  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{14}\text{N}$  and  $^{31}\text{P}$  study of aligned multilayers of 10:1 lipid:protein dispersions of DMPC, dilaurylPC (DLPC) or ditetradecylPC (DtDPC) above  $T_c$  was undertaken [33].  $^1\text{H}$  and  $^{31}\text{P}$  CSA similar to those of the pure PCs were observed, indicating that melittin does not induce any disorder in the bilayers themselves. A reduced  $^{14}\text{N}$  quadrupole splitting and  $^{13}\text{C}$  CSA for the *sn*-1 carbonyl raised the possibility of changes in conformation or motional amplitude in the lipid headgroups. An increase in  $T_1$  along with a decrease in  $T_{1\rho}$  of the methylene protons suggested that the spectral density of the lipid acyl chain motions was affected, with low-frequency motions being favored.

(-)- $\Delta^9$ -Tetrahydrocannabinol ( $\Delta^9$ -THC) is the principal psychoactive constituent of marijuana; its pharmacological activity is related to its effect on cellular membranes. The interaction of  $\Delta^9$ -THC with model phospholipid membranes was investigated by acquiring variable-temperature  $^2\text{H}$  NMR spectra of  $2\text{-}[7', 7'\text{-}^2\text{H}_2]$  DPPC and  $^{13}\text{C}$  spectra of  $2\text{-}[1'\text{-}^{13}\text{C}]$ DPPC dispersions as a function of  $\Delta^9$ -THC concentration [34]. In general, the spectra show that  $\Delta^9$ -THC lowers  $T_c$ , the temperature of the main gel-to-liquid crystal transition. With  $\Delta^9$ -THC present,  $^2\text{H}$  spectra show a more gradual change in spectral features and in the quadrupolar splitting with temperature, showing that the phase transition is broadened with respect to that of pure DPPC. However,  $^{13}\text{C}$  spectra imply the presence of both gel and fluid conformations over a narrower range of temperatures in  $\Delta^9$ -THC containing dispersions. This can be understood by realizing that the  $^2\text{H}$  spectra monitor the melting of the acyl chains which occur at the main transition temperature, whereas  $^{13}\text{C}$  spectra monitor changes in headgroup conformation which now appear to be connected with the pretransition. This study of the effects of  $\Delta^9$ -THC on DPPC allowed these two effects to be separated for the first time, as the amphipathic drug molecule interacts with the acyl chains and the headgroups in different ways, affecting their structure and dynamics to a different extent.

### 3.4. Structure of membrane proteins

Because of the problems involved in growing well-ordered crystals of large hydrophobic molecules, high-resolution structures of integral membrane proteins are difficult to obtain by X-ray diffraction. Solution NMR relies on rapid isotropic molecular motion, so it is not suitable for membrane systems in which proteins have restricted motion. Solid-state NMR has the potential for obtaining internuclear distances and orientations that can be translated into molecular structure. For membrane proteins, this has usually involved isotopic labeling combined with macroscopic orientation or MAS. In the orientation-dependent approach, the determination of torsional angles in the polypeptide backbone leads to protein structures from orientation constraints; the distance-dependent approach uses distances determined from dipolar couplings to constrain the structure [35].

#### 3.4.1. Gramicidin

Gramicidin A (gA) is a 15 amino acid long peptide which dimerizes to form an ion-transporting channel through lipid bilayer membranes. Smith et al. synthesized gA analogs, each  $^{13}\text{C}$ -labeled at a different peptide carbonyl and incorporated them into oriented DMPC multilayers [36].  $^{13}\text{C}$  spectra recorded as a function of the angle between the bilayer normal and the direction of the applied field were recorded. Simulations of the spectra allowed the determination of the orientation of each labeled carbonyl group, and the nature of its motion. The CSAs from all labeled regions were consistent with carbonyl bond orientations expected for a  $\beta^{6.3}$  helix undergoing rapid rotation about the bilayer normal.

Prosser et al. used  $^2\text{H}$  spectroscopy to study the channel conformation of gA  $^2\text{H}$ -labeled at the  $\alpha$ -carbon of residue 3, 4, 5, 10, 12, or 14 in unoriented DMPC multilayers [37]. The quadrupolar splittings yield orientational constraints on the C– $^2\text{H}$  bonds. The best fit to experimental results was given by a relaxed right-handed  $\beta^{6.3}$  helix whose axis of motional averaging is the helix axis, and no evidence for a flexibility gradient along the peptide backbone was observed. Killian et al. examined  $^2\text{H}$  NMR lineshapes and quadrupolar splittings from  $\text{Val}_1(\text{d}_8)$  gA in oriented DMPC bilayers as a function of angle with respect to the applied field. Spectra were consistent with a single-stranded  $\beta^{6.3}$  helical backbone, but not with double-helical models. The unusual side-chain orientation was explained in terms of channel-stabilizing hydrophobic interactions [38].

All of the above studies involved comparison of observed spectra with possible models of gA to obtain structural information about backbone or side-chain orientation. A recent publication by Cross and coworkers is the culmination of a series of carefully designed experiments that allowed the elucidation of a high-resolution dimeric structure of gA in oriented lipid bilayers without reference to models, solely on the basis of the amino acid sequence and 144 orientational constraints [39]. The backbone structure and indole side-chain conformations were determined with constraints based on  $^{15}\text{N}$  chemical shifts,  $^{15}\text{N}$ – $^1\text{H}$  and  $^{15}\text{N}$ – $^{13}\text{C}$  dipolar interactions, and  $^2\text{H}$  quadrupole interactions in aligned samples of labeled gA in DMPC bilayers [40–45]. The structure defines the helix as single-stranded, fixes the number of residues to 6 to 7 per turn, specifies the helix sense as right-handed, and identifies the hydrogen bonds. It should be re-emphasized that this is an *a priori* structure determined without reference to proposed models.

Most of the studies described so far involved gA in oriented bilayers. Two-dimensional REDOR NMR has been applied to investigate the conformation and dynamics of the Gly<sub>2</sub>–Ala<sub>3</sub>  $^{13}\text{C}$ – $^{15}\text{N}$  peptide bond in Val<sub>1</sub>–[1- $^{13}\text{C}$ ]Gly<sub>2</sub>–[ $^{15}\text{N}$ ]Ala<sub>3</sub>–gA in multilamellar dispersions in DMPC [46]. The  $^{13}\text{C}$ – $^{15}\text{N}$  dipolar coupling magnitude and sign were determined for labeled gA in a powder and in the multilamellar dispersion. These data allowed the set of possible angles for the peptide bond to be reduced to two values. This information eliminates the double-stranded helical dimers and the left-handed  $\beta^{6.3}$  helical dimer, but supports the right-handed, single-stranded  $\beta^{6.3}$  helical dimer as the structural model for gA in multilamellar dispersions.

#### 3.4.2. Phage coat proteins

Although not native membrane proteins, phage coat proteins incorporated into membranes are investigated by the same experiments. Filamentous bacteriophages consist of a strand of DNA surrounded by a protein coat which is made up of thousands of copies of a coat protein, each around 50 amino acids long. Coat proteins are assembled into the virus shell in a process that first involves incorporation into the membrane of the host cell [35]. The structure of the protein in the virus and in the membrane are both of interest because a comparison between them can lead to insights about the assembly process. Opella et al. have used  $^{15}\text{N}$  NMR to determine the structure of Fd coat protein subunits by measuring nuclear spin interaction tensors for all 49 peptide backbone planes in the molecule [47]. Dipole–dipole coupling, CSA, chemical

shifts and quadrupole interactions are used to provide constraints which yield a set of orientation angles for each peptide plane relative to its neighbors and to the main axis. From these relationships, an overall tertiary structure can be built up. Measurement of solution one- and two-dimensional spectra of uniformly or selectively  $^{15}\text{N}$ -labeled Fd coat proteins in micelles provided data required to obtain a tertiary structure for the membrane-bound protein [48]. The combined results of these two studies show that the protein undergoes a large change in tertiary structure on assembly: the membrane-bound form has two  $\alpha$ -helical segments, an amphipathic helix parallel to the bilayer plane and a membrane-spanning hydrophobic helix perpendicular to the bilayer plane, whereas the helices are collinear in the virus particles.

A combined solution and solid-state approach was also used to characterize a different protein, Pfl coat protein, in phospholipid bilayers [49]. Solid-state  $^{15}\text{N}$  chemical shifts were used to determine the orientation of the two helical segments shown to be present by solution NMR. A comparison of the structure with that in the virus particles showed this protein also contains two perpendicular helices in the membrane which are collinear in the virus coat. These data led to a model for the membrane-mediated virus assembly that uses the short N-terminal helix as an initiation site [50], providing some of the most detailed information available for such a process.

#### 3.4.3. *Bacteriorhodopsin*

Bacteriorhodopsin (bR), the only protein in the purple membrane of *Halobacterium halobium*, is a single 26 kD peptide chain folded into seven transmembrane helices surrounding a retinal chromophore linked to lysine<sub>216</sub> by formation of a Schiff base (SB). BR carries out light-driven proton transport which involves the photoisomerization of the retinal chromophore, the formation and decay of the M intermediate through deprotonation and reprotonation of the nitrogen, and protein conformational changes that alter the connectivity of the proton-transport pathway.

Retinal proteins can be studied by solution NMR if the membrane is solubilized first. The functional state of the protein in these samples is not known. Because bR is a large molecule embedded in a complex, rotational diffusion is too slow to average the anisotropy of the nuclear interactions, and solubility is low. Therefore, bR is a good candidate for solid-state NMR. A review of the applications of solution and solid-state NMR to retinal proteins has appeared [51].

Lewis et al. obtained  $^{13}\text{C}$  spectra of bR labeled at the backbone carbonyl of all 36 leucine residues in intact purple membrane and in DMPC vesicles. A rigid-lattice carbonyl pattern was observed for bR in the membrane, indicating that the backbone is immobile on the  $^{13}\text{C}$  NMR timescale [52]. In the vesicles, the spectrum was narrowed by fast rotational diffusion about an axis perpendicular to the membrane plane, with the extent of averaging of the chemical-shift tensor related to the orientation relative to the diffusion axis. A range of Euler angles relative to the diffusion axis for the leucine groups was calculated from the principal values of the rigid-lattice chemical-shift tensor and the motionally-narrowed lineshape in DMPC. Most (34 out of 36) leucines were in the transmembrane region, and simulations indicated that one model consistent with the data was composed entirely of  $\alpha$ -helices tilted  $20^\circ$  from the membrane normal. Bowers et al. studied bR which was  $^{13}\text{C}$ -labeled at the carbonyl of one of glycine, isoleucine, lysine, phenylalanine, or valine, and found that the proportion of mobile residues was the same as that in the C-terminus of the protein [53]. Conformation-dependent chemical shifts were used to assign the  $^{13}\text{C}$  signals of [3- $^{13}\text{C}$ ]Ala, [1- $^{13}\text{C}$ ]Ala, Leu, or Val-labeled bR to portions of transmembrane helices, loops, or the N- or C-terminus [54]. The conformation of the N and C-termini was found to be random-coil.

In addition to the peptide backbone conformation, the amino-acid side-chains of bR must play a role in the movement of protons from the SB to the extracellular surface, and from the intracellular surface to the SB. A  $^2\text{H}$  NMR investigation of side-chain dynamics in bR labeled at glycine, valine, leucine, serine, threonine, lysine, phenylalanine, tyrosine or tryptophan has been conducted [55]. All spectra exhibited a sharp central component superimposed on a broad powder pattern. The sharp signal was attributed to mobile amino-acid residues at the surface of the protein, whereas the rest of the residues appear to be immobile. Amino acids capable of ionization or hydrogenation can participate in the proton pathway; tyrosine and asparagine have been implicated by FTIR. McDermott et al. obtained  $^{13}\text{C}$  difference spectra (labeled vs. natural abundance) of tyrosine-labeled dark- and light-adapted bR and the M intermediate [56] to characterize the protonation state of tyrosine. No tyrosinate was detected in any of these. The CSA from sidebands of [4- $^{13}\text{C}$ ]Asp-labeled bR were used to determine the protonation state [57]. The internal residues Asp<sub>96</sub> and Asp<sub>115</sub> were found to be protonated, whereas all other asparagines were deprotonated.

Solid-state NMR studies have yielded much information about the retinal chromophore in bR, both in the ground state and in the M intermediate, in which the SB has released, but not replaced, its proton. In an early study, the chemical shift of [ $\epsilon$ - $^{15}\text{N}$ ] Lys-labeled dark-adapted bR was compared to that of model compounds to show that the SB was protonated [58]. It was later found that the downfield element of the C5 chemical shift tensor, the C8 isotropic shift and the  $T_1$  of C18 in a  $^{13}\text{C}$ -labeled retinal were consistent with the 6-*s-trans* conformation of the chromophore [59]. In an M intermediate containing retinal  $^{13}\text{C}$ -labeled at C5 and C12 to C15, the chemical shifts of C12 and C13 indicated that the intermediate has a 13-*cis* chromophore with an unprotonated SB [60]. Temperature-dependent quadrupole-echo lineshapes and  $T_1$  anisotropy were recorded for [18- $\text{CD}_3$ ] or [16,17( $\text{CD}_3$ ) $_2$ ] retinal in dark-adapted bR in order to calculate the activation energy for the 3-fold hopping of the methyl [61]. It was clear that the retinal chromophore is the 6-*s-trans* Schiff base. So, it appears that  $\text{CH}_3$  rotor motions can probe van der Waals contact between a ligand and a protein binding pocket. The downfield change in chemical shift of the [ $\epsilon$ - $^{13}\text{C}$ ]Lys $_{216}$  was used to follow the C=N isomerization from *syn* in bR $_{555}$  to *anti* in bR $_{568}$  [62]. Because the same label in the M intermediate resonates even farther upfield, the intermediate SB was demonstrated to be deprotonated and contain a C=N *anti* chromophore.

Rotational resonance has been used to measure internuclear distances of 4.1 Å for the 8, 18  $^{13}\text{C}$  dipolar-coupled pair, and an average C8 to C16 or C17 distance of 3.4 Å in bR, proving unambiguously that retinal is in the 6-*s-trans* conformation [63]. These experiments also suggest that it is also possible to measure dihedral angles of chemical groups using this technique. The distance between the  $^{13}\text{C}$ 14 and the  $^{13}\text{C}$   $\epsilon$ -carbon of Lys $_{216}$  in dark-adapted bR was measured to examine the structure of the retinal-protein linkage and its role in coupling the isomerizations of retinal to the unidirectional proton transfer (64). The observed values of 3.0 Å in bR $_{555}$  and 4.1 Å in bR $_{568}$  confirm that the 13-*cis* component of dark-adapted bR is 15 (C=N)*syn*, whereas the all-*trans* component is 15-*anti*. Application of rotational resonance to the same sites on the chromophore and the protein in the M intermediate gave a distance of 3.9 Å, corresponding to the *anti* C=N configuration [65]. Rotational resonance was therefore established as a feasible method to measure distances in a reaction intermediate.

### 3.5. Structure and dynamics of DNA

Alam and Drobny have reviewed the application of solid-state NMR to the determination of the structure and dynamics of DNA, focusing on deuterium NMR [66].

#### 3.5.1. High-molecular-weight DNA

High-resolution  $^{13}\text{C}$  studies of DNA are not common because of the interpretation problems caused by the broad lines observed. Nevertheless,  $^{13}\text{C}$  NMR has been used by Santos et al. to examine DNA sugar-ring puckering [67]. A series of model nucleosides and nucleotides were used to correlate the  $^{13}\text{C}$  chemical shift with deoxyribose ring conformation. High-resolution NMR of calf thymus DNA revealed that the observed trends are maintained in high-molecular-weight DNA. The major ring pucker in the A form was shown to be 3'-endo, whereas the chemical shifts of both B- and C-DNA indicated that they are mainly 2'-endo.

A study of biosynthetically  $^{15}\text{N}$ -labeled DNA grown by *E. coli* showed that 13 resonances could be resolved [68]. Protonated and non-protonated sites can be distinguished by the relative strengths of the  $^{15}\text{N}$ - $^1\text{H}$  dipolar coupling. A  $^{15}\text{N}$ - $^1\text{H}$  bond length of 1.13 Å for the G1 in the B form of DNA was calculated from the dipolar spectra, indicating that the hydrogen is localized on the hydrogen-bond donor nitrogen sites (G1 and T3), and not shared with acceptor nitrogen sites (C3 and A1). A later CP/MAS  $^{15}\text{N}$  spectrum exhibited 15 well-resolved lines [69]. A 2-dimensional  $^{15}\text{N}$ - $^{15}\text{N}$  spin-exchange experiment yielded different cross-peak intensities at different mixing times. Using the  $^{15}\text{N}$ - $^1\text{H}$  dipolar coupling as a selection mechanism, cross peaks between non-protonated  $^{15}\text{N}$  nuclei can be selected. This study suggests that in addition to the oligomers presently under investigation, the structure of large DNA fragments can be elucidated.

Solid-state  $^{31}\text{P}$  NMR has been used to examine the conformation and dynamics of the phosphodiester backbone of DNA. Shindo et al. measured the  $^{31}\text{P}$  CSA of oriented Li-DNA fibers in the A, B, and C forms [70]. The  $^{31}\text{P}$  lineshape was observed for different orientations of the fiber axis relative to the magnetic field, and a series of transformations was performed to relate the chemical shift tensor in the principal axis system to the tensor in the goniometer frame. A singlet was observed for perpendicular orientations, but calculations predicted a doublet would appear. This was interpreted as evidence for rotational motion



about the helix axis at a rate of greater than 2 kHz. A parallel spectrum broader than the perpendicular one for the B-form was accounted for by proposing a distribution of phosphodiester group orientations. Nall et al. obtained proton-enhanced  $^{31}\text{P}$  spectra of oriented fibers of A-form calf thymus Na-DNA [71]. In contrast to Shindo's results, the spectra agreed with simulations that assumed no motional averaging of  $^{31}\text{P}$  chemical shifts, an intrinsic linewidth of 5–10 ppm, and a regular phosphodiester backbone with a dispersion in the alignment of DNA crystallites relative to one another of 2 to 10°. This is in agreement with a variable-temperature study of  $^{31}\text{P}$  and  $^2\text{H}$  study of calf thymus DNA which indicates non-axial averaging of the  $^{31}\text{P}$  CSA tensor faster than 10 kHz [72], precluding the fast axial motion proposed by Shindo. A lack of averaging of the  $^2\text{H}$  powder pattern indicated there were also no large-amplitude motions on the nanosecond time scale. MAS has been used to determine the orientational distribution function of the  $^{31}\text{P}$  CSA tensor in oriented films of salmon sperm [73]. The backbone orientation obtained was consistent with that from fiber diffraction, as well as the data of both Shindo and Nall.

The  $^{31}\text{P}$  lineshape and spin-lattice relaxation of calf thymus DNA have been recorded as a function of hydration [74]. Motional narrowing increases with hydration level, and the transition from A to B form at 85% relative humidity is accompanied by a sudden decrease in linewidth attributed to backbone fluctuations of at least  $\pm 30^\circ$  at 10 kHz.  $^{31}\text{P}$  spectra of salmon sperm DNA at different relative humidities also indicated the presence of molecular motions in the hydrated fibers [75]. The dependence of the spectra on humidity could be modeled by invoking three modes of motion, namely conformational fluctuations, restricted rotation about a tilted axis, and rotational diffusion about the helical axis. In addition to hydrated samples,  $^{31}\text{P}$  solid-state NMR has been applied to concentrated solutions of DNA in aqueous NaCl [76]. Upon heating, the transition from the liquid-crystalline to the isotropic phase leads to the appearance of a sharp peak superimposed on the broad anisotropic resonance. By integration of the two peaks, a quantitative phase diagram describing the transitions occurring over a DNA concentration of 100–290 mg/ml from 20 to 60°C was constructed.

Vold et al. have used solid-state  $^2\text{H}$  to examine folded films of oriented B-form Li-DNA in which the purine bases were  $^2\text{H}$ -labeled at the 8-position [77]. Quadrupole-echo lineshape simulations were performed to determine static base-tilt angles and anisotropic motional amplitudes as a function of hydration and temperature. At 10 H<sub>2</sub>O

molecules per nucleotide, an average tilt angle of  $0^\circ$  with a standard deviation up to  $9^\circ$  was observed. The amplitude of the tilting motions increased from  $4^\circ$  in the driest samples to  $15^\circ$  in the wettest, with slightly larger amplitudes being observed for the twisting motion. Although the amplitude of the twisting motion is unaffected down to  $-60^\circ\text{C}$ , the tilting motion is suppressed at low temperatures. X-ray diffraction experiments suggested that, unlike Li-DNA, which is exclusively the B-form, Na-DNA was only the A-form. Using Na-DNA samples  $^2\text{H}$ -labeled similarly to the Li-DNA samples described above, Vold et al. showed that Na-DNA actually contains 57% of the B-form [78]. The remainder is the A-form, which was determined to have an average base tilt angle of  $23^\circ$ , with a standard deviation (reflecting a distribution of tilt angles and helix orientations) of  $4^\circ$ .

### 3.5.2. Synthetic oligonucleotides

The experiments described in the previous section have shown that there are internal motions in high-molecular-weight DNA whose rate and amplitude increases with increasing hydration and temperature. Synthetic oligonucleotides have the advantage that they can be labeled at specific sites so that local motions and orientations can be investigated. The oligonucleotide  $[\text{d}(\text{CGCGAATTCGCG})]_2$  is a self-complementary dodecamer which contains the *Eco*RI restriction endonuclease recognition site,  $\text{d}(\text{GATTC})$ , and has been characterized by many structural methods including X-ray crystallography and solution NMR. Drobny et al. therefore chose this molecule as the sample for a series of  $^2\text{H}$  solid-state NMR experiments that will be summarized here.

In one study, quadrupole-echo lineshapes and spin-lattice relaxation times were obtained as a function of hydration for two Na salts of the dodecamer [79]. One sample had the C8 protons of all the purines exchanged for deuterons, and a C6-deuterated thymidine was incorporated at the seventh position in the other. No large-amplitude motions were detected. At all levels of hydration, only rapid ( $>10^9\text{ Hz}$ ) small-amplitude ( $6\text{--}10^\circ$ ) libration of the bases was observed. At high hydration levels, a slower ( $10^4\text{--}10^5\text{ Hz}$ ) small-amplitude motion was discovered. The slow motion was attributed to a uniform twisting mode of the entire helix; its amplitude increased slightly with hydration. Collective torsional modes on the ns. time scale account for the reduced  $T_1$  at high hydration. Furanose ring dynamics in A5 and A6 were examined in the dodecamer containing 2'-deoxyadenosine labeled with  $^2\text{H}$  at the 2'' position [80]. Only a small decrease in the quadrupolar coupling con-

stant with hydration was found, precluding the occurrence of large-amplitude motions in the furanose rings up to 28 mol H<sub>2</sub>O per nucleotide. Lineshape changes which indicated motions at intermediate rates could be simulated by restricted reorientation about the helix axis which increases in rate and amplitude at high hydration levels. The dodecamer containing methyl-deuterated thymidine was also analyzed [81]. Lineshape changes were consistent with a slight increase in the amplitude of base motion with increasing hydration. The methyl-deuterated samples proved to be a good probe of the slower helix motions occurring in the range of 20  $\mu$ s, which appear at high hydration. In addition to dynamics, the deuterated samples were used to determine conformation and order in the oriented liquid-crystal phase of the dodecamer [82]. A lineshape analysis of the purine-labeled samples gave the angle of the C8 C–D bond with the helix axis to be 90°, with a spread of 20°. This agrees with the 90° orientation and spread of 15° calculated from the methyl-labeled dodecamer. The 22° (15° spread) angle of the sugar 2'' C–D bond with respect to the helix axis agrees with the expected C<sub>2</sub>-endo sugar conformation. The lineshape is consistent with a C<sub>3</sub>-endo fraction smaller than 20%.

### **3.6. Fibrous proteins**

#### **3.6.1. Collagen**

Collagen, the major protein component in connective tissue, is the most widely studied of the fibrous proteins by NMR. Its primary function is structural with collagen fibrils imparting mechanical stability to tissues. Collagen molecule consists of 3 polypeptide chains (each containing 1000 amino acids) wind around each other to form a triple helix [83], having a length of 300 nm and a diameter of 1.5 nm. Each polypeptide chain, over 1000 amino acids long, consists almost entirely of the triplet repeat, Gly–X–Y. Glycine residues occupy positions within the interior of the helix whereas the side chains of other amino acids (X and Y) are on the surface [83]. The mutual interactions among the side chains are thought to direct the assembly of collagen fibers and stabilize their structure.

*Structural studies:* Much information about the three-dimensional structure and arrangement of collagen molecules in the fibril has come from X-ray diffraction and electron microscopy [84]. However, recently the approach of selective labelling of specific sites and determining their orientation [85] to obtain the structural constraints for gradual building

of the three-dimensional structure of collagen has been attempted [86]. Saito et al. [87,88] have used  $^{13}\text{C}$  and  $^{15}\text{N}$  NMR to determine hydration/dehydration induced conformational change of collagen.

*Dynamic studies:* The assembly and stabilization of collagen molecules into fibrils is a consequence of interactions among amino acid side chains, all of which are located on the surface of the molecules. Torchia et al. [89–94] have used solid state NMR spectroscopy to elucidate these interactions and the flexibility of polypeptide backbone in collagen fibrils by studying the dynamics of specifically labeled amino acids in collagen in different tissues.

Molecular dynamics of the collagen backbone has been compared in three types of collagen fibers: (a) in reconstituted fibers of chick calvaria collagen that are not cross-linked and not mineralized; (b) in intact rat tail tendon collagen fibers that are cross-linked but not mineralized; (c) in intact rat calvaria collagen fibers that are cross-linked and mineralized [89,92]. These studies required labeling the collagen molecule with [ $1\text{-}^{13}\text{C}$ ] Gly. The results showed that the rms amplitude of the anisotropic backbone motion that occurs in reconstituted collagen fibers is slightly reduced by cross-links, whereas the amplitude of this motion is significantly reduced by mineral [92]. Multinuclear NMR studies were used to investigate the dynamical properties of the surface of the collagen molecule via the lineshape and relaxation time measurements of the side chains of different labeled amino acids [91–94]. It was found that in contrast with the results obtained for the backbone motions, side chain motions are only marginally more hindered in mineralized samples as compared with unmineralized samples [94]. These results are interpreted with reference to collagen–mineral interactions as follows [94].

Collagen fibers are composed of laterally aggregated, staggered molecules, which, when crosslinked form a strong macroscopic three dimensional network. Since the amino acid side chains are at the surface of the helical molecule, the side chain flexibility observed in the NMR experiments suggest that their interactions, responsible for the molecular stagger, involve a myriad of conformations in fluidlike domains at the molecular surface. The fluidity of the surface domains would allow collagen molecules in soft tissues to distribute stress rapidly and uniformly. The fluid domains would also facilitate mineralization by allowing rapid diffusion of ions within fibers. The backbone motions, on the other hand, provide a means of absorbing mechanical energy that is generated when tension is applied to a network of collagen fibers.

### 3.6.2. Silk

Solid state NMR has been used to characterize the molecular structure and dynamics of various silk proteins to elucidate the origin of their unique mechanical properties [95–100].

Spider dragline silk is composed of a structural protein or a family of closely-related proteins that are responsible for the unique tensile strength of fibers used for building the structural elements or spokes, of the web. Dragline silk from the Golden Orb Weaver spider, *Nephila clavipes*, is a semicrystalline polymer, and the amount, composition, and orientation of each of its phases has been established by solid state NMR spectroscopy. Dragline silk contains a 30 amino acid consensus repeat that includes a poly(alanine) region flanked by glycine-rich sequences.

Solid state  $^{13}\text{C}$  NMR chemical shifts, obtained with magic angle spinning, dipolar decoupling, and cross polarization, were used to establish that the poly(alanine) regions assume a  $\beta$ -sheet conformation [95]. This conclusion was reached on the basis of the well-known conformational sensitivity of the alanine  $\alpha$  and  $\beta$ -carbon chemical shifts [87,88]. Relaxation measurements ( $T_1$ ,  $T_2$ , and  $T_{1\rho}$ ) were used to establish that these are two motionally different populations of alanines in a ratio of 40:60%.

Solid state deuterium NMR experiments on oriented fibers of [3,3,3- $^2\text{H}$ ] alanine labeled dragline silk [96] were used to establish that about 40% of the alanine residues are very highly oriented, with their methyl groups aligned almost perfectly perpendicular to the fiber axis. This result would be expected from alanine residues in  $\beta$ -sheet conformations, if the  $\beta$ -sheets were parallel to the fiber axis. The other 60% of the alanines are also oriented with their methyl groups perpendicular to the fiber axis, but their orientation distribution function was much, much broader, approximately  $\pm 60^\circ$ . Neither of these populations of alanines underwent changes in orientation upon wetting of the fibers, suggesting that they are unaffected by plasticization by water and therefore both reside in crystalline domains.

Silk fibroin, another structural protein, is a large protein consisting mainly (about 70%) of a repeating sequence of six residues (Gly–Ala–Gly–Ala–Gly–Ser) $_n$ . Saito et al. [97] have characterized the  $\beta$  sheet conformation of silk II samples from *B. mori* fibroin and the  $\alpha$  helix form of *P. cynthia ricini* by comparing the  $^{13}\text{C}$  chemical shift of Gly, Ala, and Ser residues of the proteins with those of model peptides. Cross et al. [100] have incorporated  $^{15}\text{N}$ -Gly into natural silk fibroin to elucidate high resolution structural information for the glycine sites within the protein by analyzing the  $^{15}\text{N}$  chemical shift line shapes obtained from

oriented fibers placed parallel and perpendicular to the applied magnetic field. Their result showed good agreement with currently accepted structural models proposed by diffraction studies.

### 3.7. Ligand-protein complex

The knowledge of the bound conformation of a ligand to a particular receptor is quite important for designing potential therapeutic agents. There are many interesting ligand/receptor complexes where structural studies can not be carried out by the standard X-ray diffraction because of the lack of crystalline samples. Also because of their size, these complexes are not amenable to solution NMR studies. Recently developed solid state NMR methods can potentially be used to provide the needed structural information. A nice example of this type of structural study has recently been reported in which Schafer et al. [101] have measured internuclear distances using REDOR NMR [13], in a specifically labeled inhibitor, Cbz-Gly<sup>P</sup>-[1-<sup>13</sup>C]Leu-[<sup>15</sup>N,2-<sup>13</sup>C]Ala(ZG<sup>P</sup>LA), bound to thermolysin (a 34.6 kDa zinc endoprotease from *Bacillus thermoproteolyticus*). Their measured values of P-C', P-N, and P-C distances in this complex were consistent with those observed by X-ray diffraction in other comparable thermolysin-inhibitor complexes.

## 4. CONCLUSION

Solid state NMR provides a powerful tool for determining the structure and dynamics of many biological macromolecules not amenable (a) to high resolution solution NMR studies because of restricted dynamic properties due to their size or (b) to X-ray diffraction because of the unavailability of single crystal samples. We have outlined the nuclear interactions present in solid samples and the techniques used to obtain high resolution spectra from these samples. Since internuclear interactions are orientation and distance dependent the techniques to exploit these interactions to provide structural and dynamic information are also presented. Selected examples of the application of solid state NMR studies to questions of biomedical interest are shown.

There are increasing number of biologically important macromolecules which can not be studied either by high resolution solution NMR or X-ray diffraction techniques. Historically solid state NMR experiments have been performed mostly on home-made instruments because these experiments are technically extremely demanding. However in recent years the trend is changing because of the availability of improved

commercial instruments. It is anticipated that the availability of improved commercial solid state NMR spectrometer and the drive to understand the structural features of larger macromolecules such as membrane proteins, structural proteins and protein complexes, particularly those that are not amenable to more traditional methods, solid state NMR will increasingly become more common tool among scientists interested in biological problems.

## REFERENCES

1. C.P. Slichter, *Principles of Magnetic Resonance*, 3rd. edn. Springer-Verlag, New York, 1990.
2. M. Mehring, *Principles of High Resolution NMR in Solids*, 2nd. edn. Springer-Verlag, New York, 1983.
3. C.A. Fyfe, *Solid State NMR for Chemists*. C.F.C. Press, Guelph, Ontario, Canada, 1983.
4. J. Schaefer and E.O. Stejskal, *J. Am. Chem. Soc.*, 98 (1976) 1031
5. E.R. Andrew, A. Bradbury and R.G. Eades, *Nature*, 182 (1958) 1659.
6. I.J. Lowe, *Phys. Rev. Lett.*, 2 (1959) 285.
7. L.W. Jelinski, unpublished data.
8. A. Pines, M.G. Gibby and J.S. Waugh, *J. Chem. Phys.*, 59 (1973) 569.
9. S. Hartmann and E.L. Hahn, *Phys. Rev.*, 128 (1962) 2042.
10. G.A. Gray and H.D.W. Hill, *Ind. Res. Dev.*, 186 (1980).
11. J.S. Waugh, L.M. Huber and U. Haeberlen, *Phys. Rev. Lett.*, 20 (1968) 180.
12. D.P. Raleigh, F. Creuzet, S.K. Das Gupta, M.H. Levitt and R.G. Griffin, *J. Am. Chem. Soc.*, 111 (1989) 4502.
13. T. Gullion and J. Schaefer, *J. Magn. Res.*, 81 (1989) 196.
14. A.H. Roufosse, W.P. Aue, J.E. Roberts, M.J. Glimcher and R.G. Griffin, *Biochemistry*, 23 (1984) 6115.
15. X. Marchandise, P. Belgrand and A.-P. Legrand, *Magn. Reson. Med.*, 28 (1992) 1.
16. J.E. Roberts, L.C. Bonar, R.G. Griffin and M.J. Glimcher, *Calcif. Tissue Int.*, 50 (1992) 42.
17. A. Seelig and J. Seelig, *Biochemistry*, 13 (1974) 4839.
18. J.H. Davis, *Biophys. J.*, 27 (1979) 339.
19. R.A. Haberkorn, J. Herzfeld and R.G. Griffin, *J. Am. Chem. Soc.*, 100 (1978) 1296.
20. R.J. Wittebort, A. Blume, T.-H. Huang, S.K. Das Gupta and R.G. Griffin, *Biochemistry*, 21 (1982) 3487.
21. A. Blume, D.M. Rice, R.J. Wittebort and R.G. Griffin, *Biochemistry*, 21 (1982) 6220.

22. B.A. Lewis, S.K. Das Gupta and R.G. Griffin, *Biochemistry*, 23 (1984) 1988.
23. C.W.B. Lee, J.S. Waugh and R.G. Griffin, *Biochemistry*, 25 (1986) 37372.
24. C.W.B. Lee, S.K. Das Gupta, J. Mattai, G.G. Shipley, O.H. Abdel-Mageed, A. Makriyannis and R.G. Griffin, *Biochemistry*, 28 (1989) 5000.
25. J. Forbes, J. Bowers, X. Shan, L. Moran, E. Oldfield and M.A. Moscarello *J. Chem. Soc., Faraday Trans. 1*, 84 (1988) 3821.
26. C.W.B. Lee and R.G. Griffin, *Biophys. J.*, 55 (1989) 355.
27. M. Bloom and E. Sternin, *Biochemistry*, 26 (1987) 2101.
28. M. Auger, I.C.P. Smith and H.C. Jarrell, *Biophys. J.*, 59 (1991) 31.
29. D.B. Fenske and H.C. Jarrell, *Biophys. J.*, 59 (1991) 55.
30. C.R. Sanders II, *Biophys. J.*, 64 (1993) 171.
31. S.O. Smith, J. Hamilton, A. Salmon and B.J. Bormann, *Biochemistry*, 33 (1994) 6327.
32. E.J. Dufourcq, I.C.P. Smith and J. Dufourcq, *Biochemistry*, 25 (1986) 6448.
33. R. Smith, F. Separovic, F.C. Bennett and B.A. Cornell, *Biophys. J.*, 63 (1992) 469.
34. A. Makriyannis, D.-P. Yang, R.G. Griffin and S.K. Das Gupta, *Biochim. Biophys. Acta*, 1028 (1990) 31.
35. S.O. Smith and O.B. Peersen, *Annu. Rev. Biophys. Biomol. Struct.*, 21 (1992) 25.
36. R. Smith, D.E. Thomas, F. Separovic, A.R. Atkins and B.A. Cornell, *Biophys. J.*, 56 (1989) 307.
37. R.S. Prosser, J.H. Davis, F.W. Dahlquist and M.A. Lindorfer, *Biochemistry*, 30 (1991) 4687.
38. J.A. Killian, M.J. Taylor and R.E. Koeppe II, *Biochemistry*, 31 (1992) 11283.
39. R.R. Ketchum, W. Hu and T.A. Cross, *Science*, 261 (1993) 1457.
40. L.K. Nicholson and T.A. Cross, *Biochemistry*, 28 (1989) 9379.
41. W. Mai, W. Hu, C. Wang and T.A. Cross, *Protein Sci.*, 2 (1993) 532.
42. G.B. Fields, C.G. Fields, J. Petefish, H.E. Van Wart and T.A. Cross *Proc. Natl. Acad. Sci. USA*, 85 (1988) 1384.
43. W. Hu, K.-C. Lee and T.A. Cross, *Biochemistry*, 32 (1993) 7035.
44. P.V. LoGrasso, L.K. Nicholson and T.A. Cross, *J. Am. Chem. Soc.*, 111 (1989) 1910.
45. Q. Teng, L.K. Nicholson and T.A. Cross, *J. Mol. Biol.*, 218 (1991) 607.
46. A.H. Hing and J. Schaefer, *Biochemistry*, 32 (1993) 7593.
47. S.J. Opella, P.L. Stewart and K.G. Valentine *Quart. Rev. Biophys.*, 19 (1987) 7.
48. P.A. McDonnell, K. Shon, Y. Kim and S.J. Opella, *J. Mol. Biol.*, 233 (1993) 447.
49. K.-J. Shon, Y. Kim, L.A. Colnago and S.J. Opella, *Science*, 252 (1991) 1303.
50. R. Nambudripad, W. Stark, S.J. Opella and L. Makowski, *Science*, 252



- (1991) 1305.
51. L. Zheng and J. Herzfeld, *J. Bioenerget. Biomembr.*, 24 (1992) 139.
  52. B.A. Lewis, G.S. Harbison, J. Herzfeld and R.G. Griffin, *Biochemistry*, 24 (1985) 4671.
  53. J.L. Bowers and E. Oldfield, *Biochemistry*, 27 (1988) 5156.
  54. S. Tuzi, A. Naito and H. Saito, *Eur. J. Biochem.*, 218 (1993) 837.
  55. M.A. Keniry, H.S. Gutowsky and E. Oldfield, *Nature*, 307 (1984) 383.
  56. A.E. McDermott, L.K. Thompson, C. Winkel, M.R. Farrar, S. Pelletier, J. Lugtenburg, J. Herzfeld and R.G. Griffin, *Biochemistry*, 30 (1991) 83661.
  57. G. Metz, F. Siebert and M. Engelhard, *Biochemistry*, 31 (1992) 455.
  58. G.S. Harbison, J. Herzfeld and R.G. Griffin, *Biochemistry*, 22 (1983) 1.
  59. G.S. Harbison, S.O. Smith, J.A. Pardo, J.M.L. Courtin, J. Lugtenburg, J. Herzfeld, R.A. Mathies and R.G. Griffin, *Biochemistry*, 24 (1985) 69552.
  60. S.O. Smith, H.J. Courtin, E. van den Berg, C. Winkel, J. Lugtenburg, J. Herzfeld and R.G. Griffin, *Biochemistry*, 28 (1989) 237.
  61. V. Copié, A.E. McDermott, K. Beshah, J.C. Williams, M. Spijker-Assink, R. Gebhard, J. Lugtenburg, J. Herzfeld and R.G. Griffin, *Biochemistry*, 33 (1994) 3280.
  62. M.R. Farrar, K.V. Lakshmi, S.O. Smith, R.S. Brown, J. Raap, J. Lugtenburg, R.G. Griffin and J. Herzfeld, *Biophys. J.*, 65 (1993) 310.
  63. A.E. McDermott, F. Creuzet, R. Gebhard, K. van der Hoef, M.H. Levitt, J. Herzfeld, J. Lugtenburg and R.G. Griffin, *Biochemistry*, 33 (1994) 6129.
  64. L.K. Thompson, A.E. McDermott, J. Raap, C.M. Van der Wielen, J. Lugtenburg, J. Herzfeld and R.G. Griffin, *Biochemistry*, 31 (1992) 7931.
  65. K.V. Lakshmi, M. Auger, J. Raap, J. Lugtenburg, R.G. Griffin and J. Herzfeld, *J. Am. Chem. Soc.*, 115 (1993) 8515.
  66. T.M. Alam and G.P. Drobny, *Chem. Rev.*, 91 (1991) 1545.
  67. R.A. Santos, P. Tang and G.S. Harbison, *Biochemistry*, 28 (1989) 93728.
  68. T.A. Cross, J.A. DiVerdi and S.J. Opella, *J. Am. Chem. Soc.*, 104 (1982) 1759.
  69. S.J. Opella and K.M. Morden, *Dynamic Properties of Biomolecular Assemblies* (S.E. Harding and J. Rowe, eds.). Royal Society of Chemistry, Cambridge, 1989, pp. 196–209.
  70. H. Shindo, J.B. Wooten, B.H. Pfeiffer and S.B. Zimmerman, *Biochemistry*, 19 (1980) 518.
  71. B.T. Nall, W.P. Rothwell, J.S. Waugh and A. Rupprecht, *Biochemistry*, 20 (1981) 1881.
  72. J.A. DiVerdi and S.J. Opella, *J. Mol. Biol.*, 149 (1981) 307.
  73. P. Tang, R.A. Santos and G.S. Harbison, *Adv. Magn. Reson.*, 13 (1990) 225.
  74. M.T. Mai, D.E. Wemmer and O. Jardetzky, *J. Am. Chem. Soc.*, 105 (1983) 7149.
  75. T. Fujiwara and H. Shindo, *Biochemistry*, 24 (1985) 896.
  76. T.E. Strzelecka and R.L. Rill, *Biopolymers*, 30 (1990) 57.

77. R. Brandes, R.R. Vold, D.R. Kearns and A. Rupprecht, *J. Mol. Biol.* 202 (1988) 321.
78. R. Brandes, R.R. Vold, D.R. Kearns and A. Rupprecht, *Biopolymers*, 27 (1988) 1159.
79. A. Kintanar, W.C. Huang, D.C. Schindele, D.E. Wemmer and G. Drobny, *Biochemistry*, 28 (1989) 282.
80. W.C. Huang, J. Orban, A. Kintanar, B.R. Reid and G.P. Drobny *J. Am. Chem. Soc.*, 112 (1990) 9059.
81. T.M. Alam and G. Drobny, *Biochemistry*, 29 (1990) 3421.
82. T.M. Alam, J. Orban and G. Drobny, *Biochemistry*, 29 (1990) 9610.
83. G.N. Ramachandran, in: *Treatise on Collagen* (G.N. Ramachandran, ed.). Academic Press, New York, 1967, Vol. 1, pp. 103–183.
84. M.J. Glimcher, in: *The Chemistry and Biology of Mineralized Connective Tissues* (A. Veis, ed.). Elsevier, New York, 1981, pp. 617–675.
85. S.J. Opella, *Ann. Rev. Phys. Chem.*, 45 (1994) 659.
86. R.J. Wittebort and A.M. Clark, *Proceedings of the XIth Meeting of the International Society of Magnetic Resonance*, 14, (1992) 1.
87. H. Saito and M. Yokoi, *J. Biochem.*, 111 (1992) 376.
88. A. Naito, S. Tuzi and H. Saito, *Eur. J. Biochem.*, 224 (1994) 729.
89. L.W. Jelinski and D.A. Torchia, *J. Mol. Biol.*, 133 (1979) 45.
90. L.W. Jelinski and D.A. Torchia, *J. Mol. Biol.*, 138 (1980) 255.
91. L.W. Jelinski, C.E. Sullivan, L.S. Batchelder and D.A. Torchia, *Biophys. J.*, 32 (1980) 515.
92. S.K. Sarkar, C.E. Sullivan and D.A. Torchia, *J. Biol. Chem.*, 258 (1983) 9762.
93. S.K. Sarkar, C.E. Sullivan and D.A. Torchia, *Biochemistry*, 24 (1985) 2348.
94. S.K. Sarkar, Y. Hiyama, C.H. Niu, P.E. Young, J.T. Greig and D.A. Torchia, *Biochemistry*, 26 (1987) 6793.
95. A. Simmons, E. Ray and L.W. Jelinski, *Macromolecules*, 27 (1994) 5235.
96. A. Simmons, C.A. Michal and L.W. Jelinski, *Science*, 271 (1996) 84.
97. H. Saito, R. Tabeta, T. Asakura, Y. Iwanaga, A. Shoji, T. Ozaki and I. Ando, *Macromolecules*, 17 (1984) 1405.
98. H. Saito, M. Ishida, M. Yokoi and T. Asakura, *Macromolecules*, 23 (1990) 83.
99. M. Ishida, T. Asakura, M. Yokoi and H. Saito, *Macromolecules*, 23 (1990) 88.
100. L.K. Nicholson, T. Asakura, M. Demura and T.A. Cross, *Biopolymers*, 33 (1993) 847.

## Subject index

- accuracy, 222
  - back calculating, 223
  - constraint violation, 223
  - convergence, 223
  - direct refinement, 229
  - NMR R-value, 227
  - precision, 223
  - RMSD value, 222
- analysis of dynamics, 297
  - $^2\text{H}$  in solid samples, 299
  - chemical exchange broadening, 298
  - conformational averaging, 299
  - detecting dynamic processes, 297
  - hydrogen exchange rates, 298
  - phosphorous chemical shift, 299
- anomeric protons, 317, 319, 321
- antiphase magnetization, 252
- applications of solid state NMR, 358
  - bacteriorhodopsin (bR), 367
  - bone, 358
  - collagen, 373
  - DNA, 367
  - gramicidin, 359
  - ligand–protein complex, 376
  - lipid bilayers, 359
  - oligonucleotides, 372
  - phage coat proteins, 366
  - silk, 375
- assignment, 118
  - automated, 118, 120, 317
  - backbone, 94
  - of hydroxyl protons, 317, 319
  - methods for carbohydrates, 318
  - sequential resonance, 118
  - sidechain, 105, 107
  - SNase, 243
  - techniques for, 243
- automation of resonance
  - assignments, 229
  - assignments, 234
  - bookkeeping, 232
  - chemical shift, 234
  - databases, 234
  - pattern recognition, 230, 231
  - simulated annealing, 230
- backbone dynamics, 267
- bond lengths
  - from neutron diffraction, 249
  - influence on relaxation, 249
- CAMELSPIN, 113
- carbohydrates (*see also* oligosaccharides)
  - assignment, 318
  - $^4\text{C}_1$  conformation, 326
  - computer modelling, 328
  - interproton distances, 325
  - monosaccharides, 316
- $^{13}\text{C}/^{15}\text{N}$ -enriched proteins, 159
- CBCA(CO)NH, 101, 105
- CFN *see* coherence flow network
- chemical exchange, 37, 248
- chemical shift anisotropy, 245
- chirality
  - floating chirality, 220
  - pro-chirality, 195
- coherence, 3
  - antiphase, 6, 9
  - INEPT, 47–51, 75
  - inphase, 6, 9
  - longitudinal order, 6
  - order, 25, 26, 45
  - selection, 10, 16, 17
  - transfer, 6, 10, 46, 132, 135

- transfer pathway, 32, 77
- transverse, 6, 9
- coherence flow network, 3
  - evolution, 6
  - flow primitive, 6
  - modifier, 6
  - sink, 6
  - source, 6
  - states, 6
  - tier, 6
- conformation space sampling, 220
- conjugate gradient minimization, 255
- constraints, 196
  - chemical shift, 206
  - chiral, 215
  - covalent, 197
  - distance, 198
  - hydrogen bond, 205
  - J-coupling, 206
  - NOE distance, 198, 201
  - precision, 201
  - relaxation matrix, 200
  - torsional angle, 203
  - violation, 223
- Correlated spectroscopy *see* COSY
- correlation function, 244
- correlation time, 244
- COSY, 52
  - CFN, 52
  - interpretation, 55
  - phase cycling, 57
  - product operators, 54
  - pulse sequence, 52
  - variants, 60
- coupling, 47
  - active, 6
  - constants, 122, 125
  - operator, 14
  - passive, 6
- cross correlation, 112, 126
  - origin of, 247
  - suppression of, 252
- cross polarization, 350
- CTP *see* coherence transfer pathway
- CYCLOPS, 18, 19, 23, 26–28, 32, 45, 73
- DANTE, 250
- density matrix, 3
- deuterium quadrupole echo
  - spectroscopy, 354
- differential isotope shift method, 317
- diffusion
  - free, in a cone, 259
  - restricted, about an axis, 259
- dipolar coupling, 349
- dipole–dipole interaction, 246
- DIPSI, 67
- DNA junction, 288
- Double Quantum Filtered COSY *see* DQFCOSY
- DQFCOSY, 61
  - CFN, 61
  - phase cycling, 64
  - product operators, 62
  - pulse sequence, 61
  - variants, 65
- dynamics, 126
- E. COSY, 65, 122
- electromagnetic field
  - generation of, 242
  - rotation of, 242
- error, determination of, 263
- Euler angles, 244
- Euler’s formula, 22
- EXORCYCLE, 28–30, 35
- EXSY, 36, 39, 40, 44, 46, 50, 68
  - CFN, 40
  - phase cycling, 44
  - product operators, 40
  - pulse sequence, 39
- extreme narrowing limit, 257
- flexibility and relaxation
  - parameters, 328

- FLOPSY, 67
- fluorine-19, 319
- Fourier transform, 18, 38, 39, 42, 43, 51
- free induction decay (FID), 245
- fucose, 316, 320
- galactose, 316, 320
- Gaussian, 255
- glucose, 316, 321
- glycoconjugates, 329
  - glycopeptides, 329
  - glycopeptidolipids, 333
  - lipooligosaccharides, 333
- glycoproteins
  - glycosylation, 314
  - post-translational modification, 317
- gradients, 133
  - gradient recalled echoes, 34
  - molecular diffusion, 34
  - poor man's gradient, 323
  - pulsed field, 32, 116, 133, 138, 321
  - pulsed field gradient and spin echo sequence, 35
  - radiofrequency, 323
  - solvent suppression, 317
- gyromagnetic ratio, 246
- heteronuclear chemical shift
  - correlation, 52, 70
  - semiclassical description, 47
- heteronuclear multiple quantum correlation *see* HMQC
- heteronuclear single quantum correlation *see* HSQC
- heteronuclei, 242
- HIV, 314
- HMQC, 70, 90
  - CFN, 70
  - phase cycling, 72
  - product operators, 71
  - pulse sequence, 70
  - variants, 75
- HOHAHA *see* TOCSY
- homonuclear chemical shift
  - correlation, 52
  - COSY, 52
  - DQFCOSY, 61
  - NOESY, 68
  - TOCSY, 65
- HSQC, 74, 90, 109, 135
  - CFN, 75, 77
  - phase cycling, 77
  - product operators, 76
  - pulse sequence, 75
  - variants, 80
- hydroxyl protons, 317
- hydroxymethyl group, 318
- iduronate, 326
- INEPT, 47, 93, 131, 250
- inversion of magnetization, 31
  - phase cycle, 31
  - pulse sequence, 31
- isotope filtering experiments
- isotope labeling, 86, 89, 288
  - $^{13}\text{C}$ ,  $^{15}\text{N}$ , 86
  - deuterium, 126, 139
- isotopic enrichment methodologies, 164
  - bacteria, selective enrichment, 167
  - bacteria, uniform enrichment, 167
  - level of enrichment, 164, 175
  - mammalian, CHO cells, 173
  - mammalian, hybridoma cells, 172
  - non-biosynthetic, 174
  - pattern of enrichment, 164
  - photosynthetic bacteria, 165
  - yeast, 171
- isotropic mixing *see* TOCSY
- jump, 2-site, 260

Karplus relations, 318, 325, 326

Larmor frequency, 248

Lipari and Szabo formalism, 256

longitudinal relaxation time, 247

Lorentzian, 245

magic angle spinning, 348

magnetization transfer *see* EXSY

magnitude COSY, 46

mannose, 316, 320

mass spectrometry, 314

maximum entropy method, 326

membranes: model systems, 334

metric geometry (distance geometry), 209

- metric matrix, 210

- metrization, 210

- tetrangle smoothing, 209

model free approach, 127

models, structure calculation, 294

- distance geometry, 294

- molecular dynamics, 294

- molecular mechanics, 294

- simulated annealing, 294

molecular mechanics, 211

- DIANA, 212

- force field, 213

- molecular dynamics, 211

- Monte Carlo method, 211

- relaxation matrix, 216

- simulated annealing, 211

- time-averaged, 216

- torsion space approach, 212

Monte Carlo approach, 255

multi-dimensional NMR methods, 87–90, 317

- $^1\text{H}$ - $^{31}\text{P}$  correlation, 318

- COSY, 319

- COSY-ROESY, 324

- DQCOSY, 318

- HMBC, 320

- HMQC, 318

- HMQC-HOHAHA (TOCSY), 320

- HMQC-NOESY, 324

- HOHAHA-HOHAHA-COSY, 324

- NOESY-HOHAHA, 324

- ROESY, 326

- ROESY-TOCSY, 324

- streamlining, 320

- TOCSY/HOHAHA, 319

- TOCSY-NOESY, 325

- TOCSY-ROESY, 324

- TQF-COSY, 319

multiple pulse line narrowing, 353

N-acetylglucosamine, 314, 316

NMR detector, 17

NMR sensitivity, 314

NMR spectroscopy of isotopically enriched proteins, 175

NOE, 247

NOESY, 68

- 4D-NOESY, 87

- DNA, 285

- EXSY, 68

- $^{15}\text{N}$ -correlated NOESY, 90, 120, 121

- NOE assignments, 112, 120

- product operators, 68

- pulse sequence, 68

- RNA, 287

nonlinear least square fitting, 263

nuclear interactions in solid state, 346

- chemical shift anisotropy (CSA), 347

- dipolar, 346

- scalar, 347

- Zeeman, 346

nuclear Overhauser effect *see* NOESY

nuclear spin interactions, 244

- chemical shift anisotropy, 245

- dipolar, 245

nucleic acids, 281

- RNA, 108, 109

- oligosaccharides
  - chemical modification, 314
  - clustered, 333
  - conformation, 314
  - flexibility, 328
  - functions, 314
  - interactions, 314
  - isotope labelling, 335
- operators, 10
  - chemical shift, 10
  - lowering, 19
  - 90° pulse, 20
  - 180° pulse, 12
  - raising, 19
  - unit, 10–12
  - weak scalar coupling, 14
- order parameters, 329
  - dependence on bond length, 249
  - interpretation of, 258
  - values, 249
- P. E. COSY, 65
- PFG *see* pulsed field gradient
- phase cycling
  - automatic baseline compensation, 24
- phase cycling, 16, 132
  - COSY, 59
  - CYCLOPS, 18, 19, 23, 26
  - DQFCOSY, 68
  - EXORCYCLE, 28, 35, 79
  - EXSY, 44, 46
  - HMQC, 72
  - HSQC, 77
  - rules for, 23
  - systematic noise reduction, 24
- phase sensitive COSY, 56
- phase shift, 12
  - data routing, 22
  - effect on coherence orders, 32
  - RF pulse, 20
  - sub 90°, 27
- product operators, 3
- pseudoatom, 194
- pulse sequences
  - design, 127
  - for AX spin systems, 250
  - for AX3 spin systems, 252
- pulsed field gradient *see* gradients
- quadrature images, 16, 26
  - CYCLOPS, 18
- quadrupolar echo, 354
- random motion
  - mathematical description, 244
  - of proteins, 244
- relaxation mechanisms, 245
- relaxation parameters, 247
  - equations for, 247
  - measurements of, 250
  - sensitivity of, 248
  - values, 249
- resonance assignments, 188, 190, 281
  - 3D homonuclear NMR
  - correlation spectroscopy (COSY, TOCSY), 282
  - heteronuclear correlation, 283
  - heteronuclear, 193
  - stereospecific proton, 194
- ROESY, 113
- rotational resonance, 357
- rotational-echo double resonance (REDOR), 357
- rotations, 10
  - composite, 10
  - of electromagnetic fields, 242
  - equivalent, 11, 12
  - matrix, Wigner, 244
  - three-fold, 259
- scalar coupling constants, 324, 325
  - $^{13}\text{C}$ – $^{13}\text{C}$ , 326
  - $J_{\text{CH}}$ , 325
  - $J_{\text{COCH}}$ , 326
  - $J_{\text{HH}}$ , 325

- secondary structure, 195
  - chemical shift, 197
  - disulfide bonding, 197
  - Karplus relationship, 189
  - slowly exchanging, 196
- SEDUCE, 104
- selective excitation, 318, 321
- semi-classical description, 3
- sequential assignments, 191, 195, 283
  - amino protons, 283
  - aromatic protons, 283
  - imino protons, 283
  - J-coupling, 225
  - NOE, cross relaxation, 229
  - spin system, 232
- shaped pulses, 321
- sialic acid/N-acetylneuraminic acid, 316
- side chain dynamics, 270
  - Alanine, 261
  - Leucine, 261
- solid state NMR, 345
  - applications, 346
  - fundamentals, 346
- spectral density function, 260
- spin-echo sequence, 11
  - isolated spin system, 11
  - two-spin system — no 180° pulses, 16
  - two-spin system — selective 180° I pulse, 12
  - two-spin system — selective 180° S pulse, 15
  - two-spin system — simultaneous 180° I and S pulses, 15
- stability, 256
- staphylococcal nuclease (SNase), 261
- starting conformation, 189
  - minimization, 189
  - optimization, 189
  - structural convergence, 190
- STATES-TPPI, 98, 132
- stereo-assignment, 194, 195
- stereospecific assignments, 219
- strong coupling, 327
- structural reporter, 317
- structure determination, 290
  - coupling constants, 290
  - Karplus relationship, 290
  - NMR R factor, 293
  - NOE, 291
  - pseudorotation, 291
  - relaxation matrix, 292
- structure optimization, 217
  - annealing, 220
  - direct refinement, 217
  - four-dimensional dynamics or minimization, 219
  - minimization, 219
  - torsion space optimization, 217
- structures, examples of, 300
  - 5S ribosomal RNA, 304
  - complex-drug, protein, 306
  - DNA/RNA, 300
  - duplex oligomers, 304
  - hairpin loops, 301
  - hammerhead RNAs, 304
  - mismatched basepairs, 305
  - narrow minor groove, 300
  - pseudoknot, 303
  - RNA tetraloops, 302
  - telomers, 305
  - tetraplex, 305, 306
  - triple strands, 305
- T1, longitudinal relaxation time, 247
- T2, transverse relaxation time, 247
- three-D- and 4D- experiments, triple resonance experiments, 94
  - (H)CNH-TOCSY, 96
  - HA(CO)CANH, 92
  - HBHA(CO)NH, 95
  - HBHANH, 101
  - HCC(CO)NNH, 97
  - HCCH-COSY, 106



- HCCH-TOCSY, 106
- HCCH-TOCSY, 106
- HCP, 110
- HN(CA)HA, 92
- HN(CO)CAHA, 92
- HN(COCA)NH, 93
- HNCA, 123
- HNCACB, 92, 118, 131
- HNCACHA, 92
- HNHA, 125
- HNHA(Gly), 96
- HNHB, 122

through-space connectivities, 112, 120, 121

TOCSY, 65, 93, 107, 109, 121, 131

- CFN, 66
- product operators, 67
- pulse sequence, 66

torque, 242

TPPI, 250

transverse relaxation time, 247

two-dimensional NMR, 36

- axial peaks, 43
- cross peaks, 43
- diagonal peaks, 43
- evolution period, 36
- EXSY, 36
- frequency versus time, 38
- mixing period, 41
- preparation period, 36
- time versus time, 39

vector model, 3, 82

volume integral, 198

- integration, 198
- lineshape, 200

WALTZ, 98, 250